DESIGN OF HIGHLY DISTRIBUTED BIOFUEL PRODUCTION SYSTEMS

A Thesis
Presented to
The Academic Faculty

by

Dexin Luo

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2011

DESIGN OF HIGHLY DISTRIBUTED BIOFUEL PRODUCTION SYSTEMS

Approved by:

Professor Valerie M. Thomas, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Matthew J. Realff, Advisor
Department of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Professor Ronald R. Chance
Executive Vice President, Engineering,
Algenol Biofuels Inc
Department of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Professor David M. Goldsman
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Chen Zhou
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: October 25, 2011

To my mother, Sun Yaling

ACKNOWLEDGEMENTS

First and foremost I would like to thank my adviser, Professor Valerie M. Thomas, for her exceptional patience in giving me guidance in research and beyond, and for always being open to my ideas. She was always supportive in my career development as well as in difficult times, which shows her sincere interest in my success. I would like to express my sincere gratitude to my co-adviser, Professor Matthew J. Realff, for being exceptionally generous with his time in giving me great insights and tremendous inspiration. I have learned a great deal from both of them.

During my Ph.D. at Georgia Tech, I have had the chance to collaborate with many outstanding researchers and practitioners, to whom I am deeply thankful. Drs. Ronald R. Chance, Ben McCool, and Pradeep Sharma have provided us with valuable inputs on algae-based fuel production systems, which enabled us to extensively test some of the methods developed in this dissertation. Dr. Zushou Hu has been very helpful in providing us with and helping me understand chemical process modeling, which is essential to the work in Chapter 2. I have enjoyed working together with Dong Gu Choi on several biofuel lifecycle assessment projects. Our collaboration and discussion have enriched my understanding of different biofuel production systems. I also appreciate the opportunity to have worked with Professor Kobi Abayomi on statistical evaluation of the effect of ethanol on US corn production, which is an important part of my research experience.

I would like to thank Professor David Goldsman and Professor Chen Zhou for serving on my committee and providing valuable suggestions for my thesis. And I would also like to thank my research group for their inputs to my work over the years: Seth Borin, Todd Levin, Soheil Shayegh, Nathaniel Tindall and especially Adaora Okwo, who was always fast in providing me with information and references.

Dr. Gary Parker was very supportive when I needed help during my stay in Atlanta. Without his support, my study might have been jeopardized.

My parents always believed in me, encouraged me to follow my passions and never push any view onto me except constantly reminding me to be happy, eat healthy and exercise. Their love and encouragement are the foundation on which this thesis was built.

During my time at Georgia Tech, I have been surrounded by many incredible friends who made this process enjoyable and fun. Some of them have also provided helpful inputs and support to my research. To name a few: Anju Bhandari, J. Antonio Carbajal, Vinod Cheriyan, Shuhua Dai and Qie He, Lauren and Matthew Lee-Crane and their families, Feng Qiu, Viktoriya Rachkova, Jin Ran, Norbert Remenyi, Alison Trindall, Ayumi Nakajima and Shuangjun Xia, Fangfang Xiao, Haiyue Yu, Jiangchuan (Ralph) Yuan, and many others.

Most importantly, however, I would like to thank my husband, Alexander Epple, without whose love and support through the years, this would have been impossible.

This material is based upon work supported by the Department of Energy under Award Number DE-FOA-0000096 and also supported by Algenol Biofuels. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Contents

List of Tables

List of Figures

SUMMARY

This thesis develops quantitative methods for evaluation and design of large-scale biofuel production systems with a particular focus on bioreactor-based fuel systems. In Chapter 2, a lifecycle assessment (LCA) method is integrated with chemical process modeling to select from different process designs the one that maximizes the energy efficiency and minimizes the environmental impact of a production system. An algae-based ethanol production technology, which is in the process of commercialization, is used as a case study. Motivated by this case study, Chapter 3 studies the selection of process designs and production capacity of highly distributed bioreactor-based fuel system from an economic perspective. Nonlinear optimization models based on net present value maximization are developed that aim at selecting the optimal capacities of production equipment for both integrated and distributed-centralized process designs on symmetric production layouts. Global sensitivity analysis based on Monte Carlo estimates is performed to show the impact of different parameters on the optimal capacity decision and the corresponding net present value. Conditional Value at Risk optimization is used to compare the optimal capacity for a risk-neutral planner versus a risk-averse decision maker. Chapter 4 studies mobile distributed processing in biofuel industry as vehicle routing problem and production equipment location with an underlying pipeline network as facility location problem with a focus on general production costs. Formulations and algorithms are developed to explore how fixed cost and concavity in the production cost increases the theoretical complexity of these problems. Appendix B, as an appendix to Chapter 3, explores how transportation cost scales with total production rate and its impact on capacity design.

Chapter I

INTRODUCTION

In light of the growing global need for renewable energy production, comprehensive evaluation of all possible energy technologies is of significant importance, especially when determining which technologies to choose for large-scale production. Renewable energy technologies are usually proposed because of their naturally replenished nature and their alleviated effect on climate change. However, technology evaluation should not only include environmental attributes such as energy use and greenhouse gas (GHG) emissions, but also the practicality of the proposed technology from an economic perspective. Design of processes and production systems plays an essential role in commercialization of emerging energy technologies, and therefore should be a significant component of the technology selection process. While production design usually focuses on cost and efficiency, it is important to incorporate the environmental factors in the early design stage of energy production processes to maximize the overall environmental performance of the product.

Although energy technologies have been studied for decades, economic analysis, environmental evaluation, and systems design are often carried out independently because of the complexity of energy production systems and different disciplines involved. Integrated approaches are needed in the literature for design and analysis of energy production systems. This thesis, inspired by an algae-based ethanol production process, focuses on large-scale bioreactor-based fuel systems for production of biofuels and other chemicals. The thesis develops and implements models that incorporate economic, physical and environmental attributes for designing and analyzing such systems.

Advances in algae-based ethanol production technologies have led to studies of different processes. Chemical and fuel industries have announced their intention to produce

ethanol and other chemicals such as polyethylene from algae. In recent years especially, vigorous initiatives on developing biofuels as alternatives to fossil fuels have been growing globally, mostly driven by policies focusing on energy security and mitigation of GHG emissions. There are, however, increasing concerns that large-scale ethanol production from corn and cellulosic materials will cause food competition and land use change. Algae-derived biofuel as the third generation biofuel is considered to be a promising alternative [139] [24]. Raceway ponds are the most commonly known production environment for algae-based biofuels. Recently, governments and companies started investing heavily on closed bioreactor systems because of their advantages such as easier contamination control and less water usage [136]. In spite of the increasing number of studies on algae-based biofuels [139] [93] [118] [164] [88] [63] [163], it is still a limited field of study [136]. Particularly, there is no known literature on system design of the large-scale algae-based production systems. This thesis starts with lifecycle analysis and process design of one algae-based biofuel and extends to comprehensive economic evaluation and production network design of bioreactor-based large scale fuel systems.

The bioreactor-base production system studied in this thesis includes not only algae systems but any large scale micro-bacteria system that is characterized by highly distributed bioreactors, intensive liquid and gas transport and aggregated process unit operations. A formal description of such systems is presented in section 3.2. The following sequence forms the basis for a biofuel production process that is used as a case study throughout this thesis: Ethanol is produced via an intracellular photosynthetic process in cyanobacteria (blue-green algae), excreted through the cell walls, collected from closed photobioreactors as a dilute ethanol-in-water solution, and purified to fuel-grade ethanol. This production process involves intensive liquid and gas transport: carbon dioxide is transported from a nearby source to each photobioreactor at the plant, the chemical mixture collected from each photobioreactor is transported to processing facilities where ethanol is separated from water, water is recycled and returned to each photobioreactor, and flue gas from the

headspace of each photobioreactor is sent to scrubbers where ethanol and carbon dioxide are recycled and oxygen is released. Hence, the liquid and gas transportation network is a major component of the production cost of such systems. There are also nutrient inputs into the photobioreactors and biomass waste as outputs of the photobioreactors.

Most biofuel production systems are slow production systems. The photosynthesis and growth of biomass is slow compared to the industrial process systems such as distillation or reaction, i.e. the point sources (biomass) produce the fuel relatively slowly compared to the ability of the next step to process it. In most biofuel systems such as corn ethanol production, biomass is collected as a feedstock and fed into industrial processes, and fuel is produced at a speed of the feeding process. Comparatively, the processes considered in this thesis does not necessarily have biomass as feedstock. For example, in the ethanol production from algae process, ethanol is produced at the speed relevant to the algae growth. There are similar production systems:

- Very small scale ethanol fermentation at the farm level, where very cheap fermentation systems (e.g. plastic bags) are used to do the fermentation at the farm. The fluid produced by the fermentation is then collected by truck and taken to a central refining site to be purified to ethanol. The biomass residual is reused at the farm as a fertilizer

- Small scale biodiesel production where reaction of triglycerides is done at small scale. The waste glycerol and methanol is trucked to a central location for purification and upgrading.

- Waste oil processing, where waste oil fractions are collected from many restaurants and then processed through a biodiesel process, similar to the one above.

- Pyrolysis oil generation, where a pyrolysis unit is used to create pyrolysis oil around various locations. The pyrolysis oil is then taken to a central location to be processed into fuel. This process has two characteristics: 1) there is a mobile processing system

which must be moved around to create pyrolysis oil, 2) there is a central location or multiple regional locations for oil production

There is few literature on production planning and cost analysis of such systems. The effect of such production process on timing of investments and the optimal scaling of the system is studied in this thesis.

This thesis studies three design stages of highly distributed bioreactor-based fuel systems: 1) the evaluation of candidate processes through LCA on their energy use and GHG emissions, 2) strategic planning of the production capacity, and 3) tactical design of the production layout. The purpose of this chapter is to provide an overview of motivations, research questions and contributions of this thesis.

Lifecycle Assessment and Process Design for Ethanol Production from Blue-green Algae

Energy production is usually of large scale and has considerable energy consumption and greenhouse gas emissions. To meet energy demand while regulating GHG emissions, governments favor processes that have low GHG emissions and high energy return on investment (EROI). Companies may be required to do lifecycle assessment on their production systems for government and investor approval and support. Chapter 2 contributes to the literature by providing a lifecycle assessment of algae-based ethanol production processes, which has not been addressed sufficiently in the existing literature. In this thesis a new ethanol production technology based on blue-green algae proposed by Algenol Biofuel Inc. is evaluated. Currently, there is a fierce debate on the attractiveness of producing ethanol from algae [35] [98]. The first part of the thesis aims at providing a complete understanding of the lifecycle energy use and GHG emissions of one algae-based biofuel. Factors that contribute most to the environmental impact of the process are identified and the potential for improving the environmental performance of the product is quantified.

Besides the traditional use of LCA for product assessment, LCA can also be used as an effective tool in process design. Process design has traditionally been focusing on cost

and efficiencies but usually fails to analyze the environmental performance of candidate processes in early design stages. LCA allows evaluation of the environmental effect of every step of a process over its entire life cycle, which helps the process designer identify potential for improving energy efficiency and reducing emissions. Using LCA to assist process design is not a new concept but is still rare in the literature [100] [4] [11] [26]. In addition, few LCA studies have addressed the issue of heat and power integration in process design even though heat recovery is essential to system efficiency and performance [100]. Chapter 2 presents a new application of LCA in process design. Detailed engineering scenarios are developed to show the advantage and disadvantage of different processes over the product life cycle.

Standard LCA [65] is combined with chemical process modeling [117] in Chapter 2 to assist process design in producing ethanol from blue-green algae. Using process modeling in LCA allows for the systematic calculation of environmental impact when changing the process design [171] [97] [29]. The goal of the study is to compare the environmental performance of candidate processes under different energy supply and technology development scenarios. The scope of the LCA is defined in figure 2, which is followed by an inventory analysis of relevant energy and material inputs and environmental releases. The potential environmental impacts associated with identified inputs and releases are then evaluated, the result of which is then interpreted to provide a basis for decision-making. Different from most LCA work in which all data are from academic literature or industry inputs, the electric energy and heat needs for a candidate ethanol separation process are obtained from chemical process modeling. Basic thermodynamics calculations are used to verify the data obtained from the process modeling. In addition, three scenarios for supplying electric energy and heat are developed and the corresponding energy use and greenhouse gas emissions are calculated. The connection of process modeling and the LCA method applied in biofuel production contributes to an important and inadequate field in literature. The use of detailed chemical process modeling improves the quality of data

used for inventory calculation and lends credibility to the LCA results.

Optimal production capacity of highly distributed bioreactor based fuel systems

Deep understanding of how production cost scales with production rate is essential to the design of large-scale production systems. The fossil fuel industry, for example, generally consists of giant refineries, because the unit cost of production decreases monotonically with increasing plant size. Most biomass utilization facilities that require low-density biomass transportation have smaller optimal capacities because of the competition between the scaling effects of transportation cost and processing cost [78] [169]. The scale economies of bioreactor-based fuel production systems such as ethanol produced from algae are not well understood in the literature. In fact, there is little research on the cost estimation of large-scale biofuel production from algae. The optimal production capacity for algae-based production systems is yet to be explored.

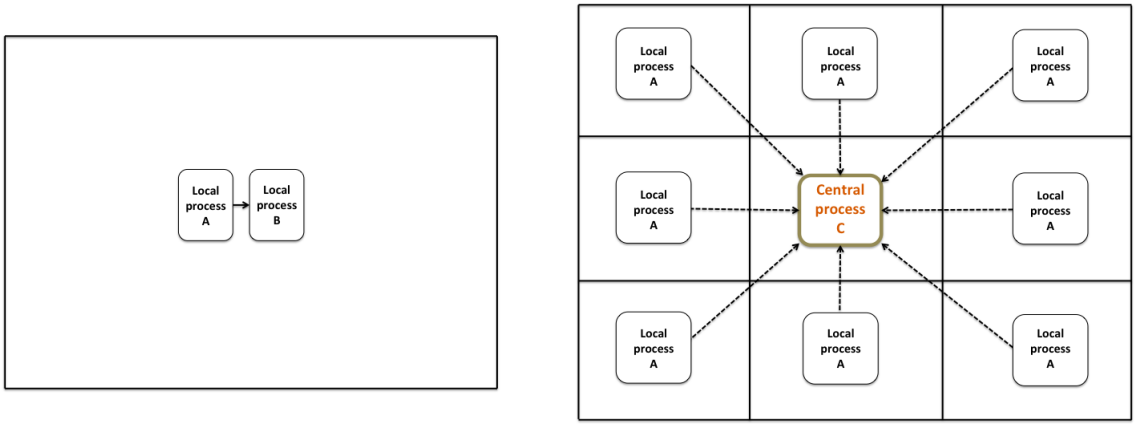


Figure 1: Layout of two different process designs studied in Chapter 3

The appropriate production scale and the potential economic viability of these systems depend on a number of uncertain factors. Chapter 3 develops nonlinear optimization models based on net present value maximization to analyze production scale and economic viability of bioreactor-based fuel systems, in which chemical mixture is collected from each bioreactor and transported to processing facilities. The ethanol production process from blue-green algae described in Chapter 1 is used as a case study. Two production

layouts are studied as shown in figure 1: 1) chemical mixture from each bioreactor is transported to a central facility to be produced into the final product by an integrated process that may contain two or more process steps; 2) chemical mixture from the bioreactors are first transported to a distributed process to be concentrated into more valuable stream and then transported to a central location to be processed into the final product. The production capacity of the plant is determined by the outlet flow of the central process. The land area of each basic production system is then determined by the number of and size of the bioreactors.

Because the scale of central process is large, the timing of production becomes an important factor. Process columns, such as distillation, requires an inlet flow rate to reach a certain percentage of its total capacity to start operation. In order to reach the inlet flow rate, corresponding number of bioreactors and distributed process columns have to be built. There is a limit on the number of bioreactors that can be built in a day. Since algae growth is not a fast process, the larger capacity of central process requires larger numbers of bioreactors to be built, which results in longer delay of operation, and hence loss of revenue. The trade off between time value of money and economies of scale contributes to an optimal capacity for the central process.

The models presented in Chapter 3 capture the scale economies of capital and operational cost of production equipment, transportation and bioreactors. By incorporation the timing of the revenue generation process and cost spending process in the decision models, this chapter shows that time value of money plays an important role in capacity sizing for bioreactor-based fuel systems. Global sensitivity analysis based on Sobol's method indicates that the top three factors affecting the optimal scale of production are the rate of biofuel production in the bioreactors, the rate of building bioreactors, and the cost of capital. The top four factors affecting financial viability are the price of ethanol, the cost of capital, the concentration of biofuel in the fluid extracted from the reactors that is sent to the

processing units, and the rate of biofuel production in the bioreactors. Conditional value-at-risk minimization shows how the optimal decision depends on the risk inclination of the decision-maker; for a risk-neutral decision maker in the case study, the optimal production scale is about 320 million gallons of ethanol per year, and about 120 million gallons per year for risk-averse decision-makers.

Formulation and Heuristic Solution for Location and Routing Problems of Some Fixed and Mobile Processing Options for Biofuels Production

The focus of Chapter 4 is two types of transportation problems applied in the biofuel industry with a focus on general production cost structure. The possible concavity in production cost typically occurs because of economies of scale in production.

The first type of problem involves locating and sizing production equipment on a given undirected complete graph. This problem is motivated by the bioreactor-based ethanol production from algae described in Chapter 2, where each bioreactor is considered as a source node and where the flow are transported through pipeline networks. Mixed integer optimization models based on bipartite graphs are commonly used in the biofuel literature to model facility location problems [22] [39] [176], allowing only connection between a source node and a potential facility site. When the transportation network is based on pipeline system, however, connection between sources should be allowed. The underlying transportation problem is different from the facility location problem on bipartite graph. In Chapter 3, the source nodes (bioreactors) are assumed to be packed on a rectangular area and located systemically around a central location. The distributed processes are assumed to be the equivalently sized, and also symmetrically located around the central process. This assumption is relaxed in Chapter 4, where the source nodes are randomly located and the number of distributed processes and their locations are to be determined. Optimization models based on facility location problem known to the operations research literature are developed. The impact of the production equipment cost structure on the problem complexity is studied. Polynomial algorithms are presented for both the case with linear production

cost and that with concave production cost.

The second type of problem studied in this chapter is motivated by the mobile distributed processing technology in the biofuel industry. Mobile processors could be moved around to each biomass source to process the raw material into more concentrated and more valuable product streams, which is then carried by these processors to the central facility to be processed into final products. This is believed to reduce transportation cost when the source nodes are highly distributed with low density material flow. The research question here is: Given the location and demand of all sources, what should be the size of the mobile processors and how many of them should be used. A subsequent question is then how should these mobile processors be routed to serve all the given sources. This type of problem has not been addressed in the biofuel literature. Chapter 4 models this problem as capacitated vehicle routing problem with certain time constraints. Continuous source such as algae-based biofuel and static source such as terrestrial biomass are differentiated with different modification to the model and different time constraints. Route first-cluster second algorithm is modified to solve the models proposed. The modified algorithms are tested on two case studies that involve, respectively, static and continuous sources.

Chapter II

LIFECYCLE ASSESSMENT AND PROCESS DESIGN FOR AN ETHANOL PRODUCTION PROCESS BASED ON BLUE-GREEN ALGAE

2.1 Problem Context

Key challenges in the production of biofuels are to limit the competition with agriculture, to reduce other land use impacts, and to keep greenhouse gas emissions low. Biofuel produced from perennial feedstocks such as switchgrass and miscanthus can have very low greenhouse gas emissions because a significant portion of the biomass is used to generate process heat and electricity, with extra electricity produced and exported to the grid. These systems can even have negative greenhouse gas emissions if the extra electricity production is counted as displacing fossil-derived electricity [135]. Perennial feedstocks can, however, affect agricultural land use, soil carbon, and ecosystems, both directly and indirectly. Managing these effects presents a significant challenge to the large-scale development of biofuels.

Biofuel can also be produced from microalgae and cyanobacteria. These require less land area than other biofuel systems, and need not use agricultural or environmentally sensitive land [163] [105]. The most commonly discussed approach to algal biofuel is to grow and harvest algae, and to process the dewatered algae to yield a biodiesel fuel. The energy and greenhouse gas impacts of this type of system have been addressed previously [93] [35].

Here we address a different type of biofuel system, in which cyanobacteria (blue-green algae) themselves make ethanol. The cyanobacterial cultures are not harvested, but are maintained for ongoing ethanol production. The key processing step involves separating

the ethanol from the sea-water growth medium. Such a no-harvest strategy mitigates the high energy costs and high water usage associated with the separation processes required for algae harvesting and fuel extraction [167] [122] [166]. In this chapter, we will explore a specific non-harvest case for ethanol production as described by Algenol Biofuels. We rely on publically available information for the basic approach [167] [166] [12], carry out engineering calculations for some of the key processes, and invoke ranges for parameters and system components that cannot otherwise be specified.

The ethanol-producing organisms are long-lived blue-green algae (genetically enhanced photoautotrophic cyanobacteria) grown in closed photobioreactors containing seawater supplemented with carbon dioxide and small amounts of nitrogen and phosphorus fertilizers. Several species of cyanobacteria, including Chlamydomonas reinhardtii, Oscillatoria limosa, Microcystis PCC7806, Cyanothece PCC 7822, Microcystis aeruginosa PCC 7806, Oscillatoria sp., and Spirulina platensis produce ethanol photosynthetically; strains can be selected for ethanol-tolerance, salt-tolerance, and pH tolerance and ethanol production can be enhanced through genetic modification [12]. The ethanol is collected as a dilute ethanol-freshwater solution from the cyanobacteria-seawater culture [167], and purified to fuel-grade. Along with 18 other technologies, including cellulosic approaches, this technology has been selected by the U.S. Department of Energy for biorefinery development at the pilot plant scale. Several processes are proposed to purify the dilute ethanol-in-water solution to fuel-grade ethanol, each of which has different environmental impacts.

The US Energy Independence and Security Act requires that cellulosic renewable fuels have lifecycle greenhouse gas emissions of less than 40 percent of petroleum-derived fuels [153]. In addition, the act provides funding to support development of renewable fuels that have lifecycle greenhouse gas emissions of less than 20% of petroleum derived fuels, making 20% the de facto goal for biofuels. The US EPA uses year 2005 gasoline as its baseline for comparison, with lifecycle greenhouse gas emissions of 91.3 g CO_2e/MJgasoline [141]. In this chapter, we calculate the lifecycle energy and greenhouse gas emissions for three

different system scenarios for three proposed ethanol production processes, using process simulations and thermodynamic calculations.

The concentration of ethanol in the liquid collected from the cyanobacteria system will strongly affect the amount of energy needed to concentrate the ethanol: the higher the initial concentration of ethanol, the less energy is needed to purify the ethanol. There is no published information on the ethanol concentration levels that can be achieved in the cyanobacteria cultures, though there will be some limiting value dependent on the organisms' tolerance. Here we evaluate the lifecycle greenhouse gas emissions as a function of initial ethanol concentration, and report the results for initial concentrations (given in weight per cent) ranging from 0.5% to 5%. This choice of ranges is somewhat arbitrary, but has been made based on the likelihood that 0.5% is too dilute for economical recovery and that 5% would have a high likelihood of economical recovery, in that the separation process could employ standard column distillation.

2.2 Product Lifecycle and Analysis Framework

Figure 2 shows the scope of our lifecycle analysis of the cyanobacterial ethanol production process. In calculating the lifecycle energy and greenhouse gas emissions, we consider the production of the cyanobacteria, including the production and disposal of the photobioreactors, mixing in the bioreactors, the disposal of the waste biomass, the production and transport of the nitrogen and phosphorus fertilizers, ethanol separation processes, ethanol transportation and distribution, and ethanol combustion in motor vehicles. For the ethanol separation processes, responsible for most of the lifecycle energy use and greenhouse gas emissions, process-based calculations have been developed, using Aspen Plus unit operations simulations as well as literature results. In addition, we use thermodynamics to provide a transparent characterization of energy requirements and resulting greenhouse gas emissions. Details are provided in the appendix A.

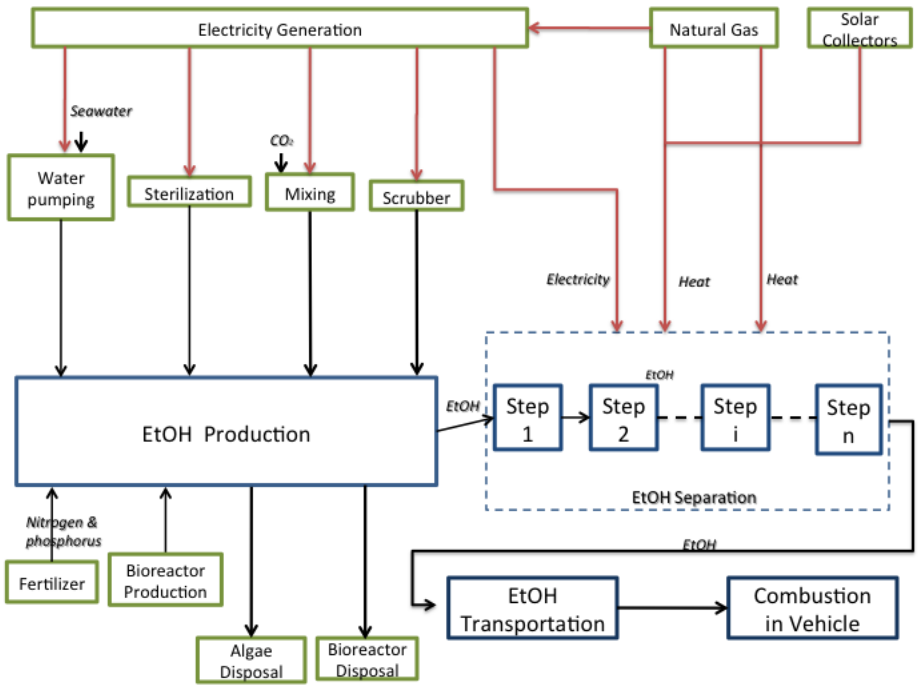


Figure 2: LCA framework

In the concept [167] [166] to be modeled here, cyanobacteria will be grown in flexible-film, polyethylene-based photobioreactors containing seawater or brackish water as the culture medium. Since additional carbon dioxide is required to support efficient algal growth, the production facility will be located near a fossil-fuel power plant or industrial source of carbon dioxide. Here the calculation is based on use of industrial CO_2, such as the by-product CO_2 from ethylene oxide production. Carbon dioxide could be dissolved in the seawater growth medium or injected into the headspace of the photobioreactor; here we develop the calculation based on introducing the CO_2 into the headspace. Nitrogen and phosphorus fertilizers will be introduced into the photobioreactors to support the initial cyanobacterial growth. At mature operation, the cyanobacteria will produce ethanol which will diffuse through the cell wall into their growth medium. There will be water and ethanol in the vapor phase inside the photobioreactors; the freshwater/ethanol mixture collected from that vapor as a chemical mixture [167] will undergo a series of separation processes

to produce 99.7% pure (fuel-grade) ethanol. Algenol aims to produce about 56,000 liters of ethanol per hectare per year using about 430 polyethylene photobioreactors per hectare, each with about 4500 liters of culture medium containing about 0.5 g/l of cyanobacterial biomass [166]. This production target is within achieved photosynthetic yields (2-4) and corresponds to 1.8% solar energy conversion efficiency for average incident sunlight energy levels in the United States [105]. However, since the system is still in the development phase, we will consider herein a range of production rates to test sensitivity to the yield assumption. The photobioreactors are modeled as having a length of 50 feet and a cylindrical diameter of 4 feet, composed of polyethylene of 0.2 mm thickness. The cyanobacteria will remain in the photobioreactors producing ethanol, and unlike other algae-to-fuel processes, will not be harvested for fuel or other purposes. It is anticipated that the photobioreactors will be emptied no more than once per year to replace the seawater, growth media and cyanobacteria.

The analysis studies three ethanol purification processes that require different energy inputs to achieve the same goal, i.e. to purify the ethanol-in-water solution to fuel grade ethanol.

- Process 1(P1): Vapor compression steam stripping (VCSS) plus vapor compression distillation (VCD) plus molecular sieve
- Process 2 (P2) Vapor compression steam stripping (VCSS) plus conventional distillation plus molecular sieve
- Process 3 (P3): Vapor compression steam stripping (VCSS) plus membrane separation

Electricity and heat are needed for the ethanol purification processes, water pumping, sterilization, mixing and scrubber. The following three energy supply scenarios are proposed for evaluation of the potential GHG emissions over the product life cycle.

- Scenario 1 (S1): U.S. grid electricity for electricity and natural gas for process heat need.
- Scenario 2 (S2): Combined heat and power (CHP) with gas turbines for electricity and process heat need with natural gas.
- Scenario 3 (S3): Combined heat and power (CHP) with gas turbines for electricity production and some process heat, provide additional heat need with solar thermal.

The methodology and analysis will be explained using a baseline case and the final results will be shown at the end of the chapter. The baseline case considers 1% initial concentration, P1 for ethanol separation and S2 for energy supply.

2.3 Analysis of Ethanol Separation Processes

2.3.1 Candidate Separation Processes

Though standard column distillation could be applied to much of the chosen range of initial ethanol concentration, the energy efficiency of that process falls off rapidly below 5% [155]. Therefore, for the first separation stage for the chosen concentration range, a unit operation based on vapor compression steam stripping (VCSS) is modeled, a highly heat integrated process that offers the potential for energy efficient separation even at low ethanol concentrations [155] [156] [133] [132]. The VCSS unit will concentrate the ethanol to a value in the 5-30% range depending on the starting value and other details of the process.

To concentrate the ethanol to 94% (at or near the azeotrope), both vapor compression distillation and conventional distillation are simulated for comparison. There are a number of choices for the final stage to fuel grade ethanol (99.7%) including molecular sieves, extractive distillation, and membrane assisted processes [155] [156]. Molecular sieve dehydration is chosen for the present study. These result in the first two chosen processes P1 and P2 for ethanol separation. The complete ethanol purification process can be achieved by membrane process (P3). Figure 3 shows the ethanol production and purification process

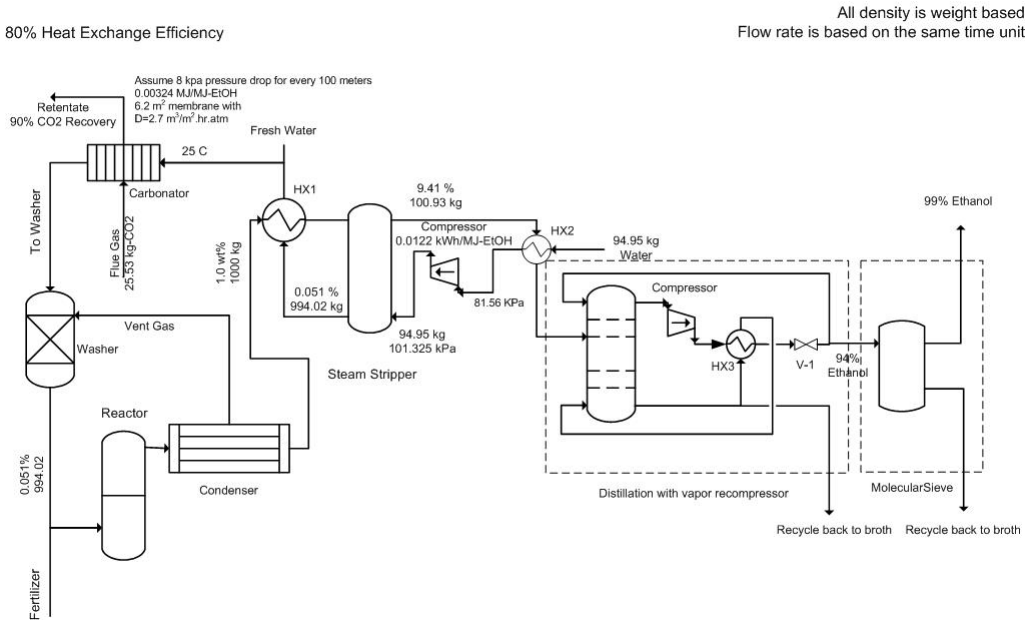


Figure 3: Process simulation diagram for cyanobacterial ethanol production. The rates are with respect to a flow of 1000 kg from the condenser. *Simulated by Zushou Hu*

for P1.

2.3.2 Model Formulation and Energy Analysis

Let E_e and E_h be, respectively, the electricity and fossil heat requirementof the ethanol separation process; and let G_e and G_h be their corresponding greenhouse gas emissions. Further, let E_j and G_j be the energy use and greenhouse gas emission of component j of the product life cycle other than the separation process, e.g. the fertilizer production. The energy use for scenarios $S1, S2$ and $S3$ are:

$$E^{S1} = E_h + \frac{E_e}{\eta} + \sum_j E_j \tag{1}$$

$$E^{S2} = E_h - E_e\theta + \frac{E_e}{\eta} + \sum_j E_j \tag{2}$$

$$E^{S3} = E^{S2} \tag{3}$$

where η is the electricity production efficiency, and θ is the steam-to-power ratio when combined heat and power is used to produce electricity. The greenhouse gas emissions for scenario $S1, S2$ and $S3$ can be calculated as follows:

$$G^{S1} = g_n E_h + g_e E_e + \sum_j G_j \tag{4}$$

$$G^{S2} = g_n (E_h - E_e \theta) + g_c E_e + \sum_j G_j \tag{5}$$

$$G^{S3} = \gamma g_n (E_h - E_e \theta) + g_c E_e + \sum_j G_j \tag{6}$$

where g_n is the GHG emission of supplying one unit of heat using natural gas, g_e is the GHG emission of supplying one unit of electricity using U.S. grid electricity, g_c is the GHG emission of supplying one unit of electricity using combined heat and power and γ is the percentage of extra heat need that is supported by natural gas. The GHG emission from solar thermal is very small and is assumed to be zero in this analysis.

E_e and E_h are simulated using chemical process modeling and verified by basic thermodynamic calculation. In the baseline scenario process heat is provided from natural gas. Ethanol production will vary with temperature and light levels [164]; we develop a time-average calculation. The vaporization and compression aspects are discussed separately below using 1% concentration to explain the calculations in detail. The final sections show the results for a range of concentrations ranging from 0.5 to 5%.

Vaporization Process

The concentration of ethanol produced by the steam-stripping column is determined by the input concentration of ethanol and the operating temperature of the column. For a 1% initial ethanol concentration, the steam-stripping column, operated at atmospheric pressure, will increase the ethanol concentration from 1 percent to 9.4 percent, and the vapor compression distillation (VCD), operated at atmospheric pressure, will raise the concentration of ethanol from 9.4% to 94% (details are provided in appendix A).

The energy required for vaporization in a steam-stripping column can be estimated from the heats of evaporation of water and ethanol:

$$E_{evap} = \sum_i m_i \Delta_v H(T)_i \tag{7}$$

where m_i and $\Delta_v H(T)_i$ are the mass flow rate and the heat of evaporation of compound i. The heat of vaporization of pure ethanol is 838 kJ/kg and the heat of vaporization of water is 2260 kJ/kg [96]. Using equation 1, the heat required for the steam-stripping process is 0.85 MJ/MJ_{EtOH}, a value that is close to our simulation result of 0.89 MJ/MJ_{EtOH}. The vapor from each steam-stripping column is condensed with a heat exchanger; the heat released from the condensation provides the heat for the steam-stripping column via a plate heat exchanger. The efficiency of this heat exchange process is an important variable in the assessment of overall energy efficiency of the VCSS system. This efficiency can be described in terms of the approach temperature for the heat transfer process, or, more transparently for our purposes, as a heat exchange efficiency. For plate heat exchangers, 80% heat recovery is achievable in practical devices, with higher recoveries possible at higher capital cost [89]. As a baseline we adopt the conservative assumption of an 80% efficient heat exchange. In that case the net heat input into the steam stripping, from our simulation results, is 0.18 MJ/MJ_{EtOH} for our 1% reference case. We also calculate the results for a more optimistic assumption of 90% efficient heat exchange. In that case, the net heat input is 0.09 MJ/MJ_{EtOH}. We use the conservative 80% assumption for the bulk of the chapter and make comparison to the 90% case at the end. The process heat requirement is driven by the initial concentration of ethanol. The energy requirement for evaporation, equation 1, can be re-expressed, per unit of ethanol for a two-component, water and ethanol, systems, as

$$
\begin{aligned}
e_{evap} &= \frac{k}{x}[x\Delta H_{EtOH} + (1-x)\Delta H_{H2O}] \\
&= \frac{k}{x}[\Delta H_{H2O} - x(\Delta H_{H2O} - \Delta H_{EtOH})] \quad (8) \\
&\approx \frac{k}{x}\Delta_{H2O} \quad (9)
\end{aligned}
$$

where x is the ethanol fraction, e_{evap} is the energy per MJ of ethanol, k is the proportionality factor, and the final approximation is for $x << 1$. This inverse relationship at small concentrations illustrates the importance of increasing the initial ethanol concentration in order to keep lifecycle energy requirements low. Improvements in the energy efficiency of the initial stage of the ethanol separation process could provide a substantial energy benefit.

Compression Processes

Steam compression is required for the stripper column and the VCD column; the compressor is electrically powered. For 1% initial concentration, process simulation (Aspen Plus) was used to derive a steam compression requirement for the steam stripper of 81.56 kPa to 101.32 kPa, as shown in Figure 3 and described in the appendix.

The work required for adiabatic compression in an open flowing system is

$$
W_{adiabatic} = \frac{\gamma nRT}{\gamma - 1}[(\frac{P_{out}}{P_{in}})^{\frac{\gamma-1}{\gamma}} - 1] \quad (10)
$$

where R is the gas constant, T is the temperature in Kelvin, P is the pressure, and γ is the adiabatic coefficient, taken to be 1.3 (details provided in SI). The adiabatic work required for the compression is 0.0058 $kWhe/MJ_{EtOH}$. Simulation using Aspen Plus yields $0.0051 kWhe/MJ_{EtOH}$ for the VCSS compression and $0.0067 kWhe/MJ_{EtOH}$ for the VCD compression, for a total of 0.0118 $kWhe/MJ_{EtOH}$. If the electricity were produced with 38% efficiency this would require 0.11 MJ/MJ_{EtOH}.

Final Purification from the azeotrope to fuel grade ethanol

For the molecular sieves stage, the total heat requirement is estimated to be 1 to 2 MJ/kg_{EtOH} [32]. We use 1.5 MJ/kg_{EtOH}, which is equivalent to heat requirement of E_{ms} =

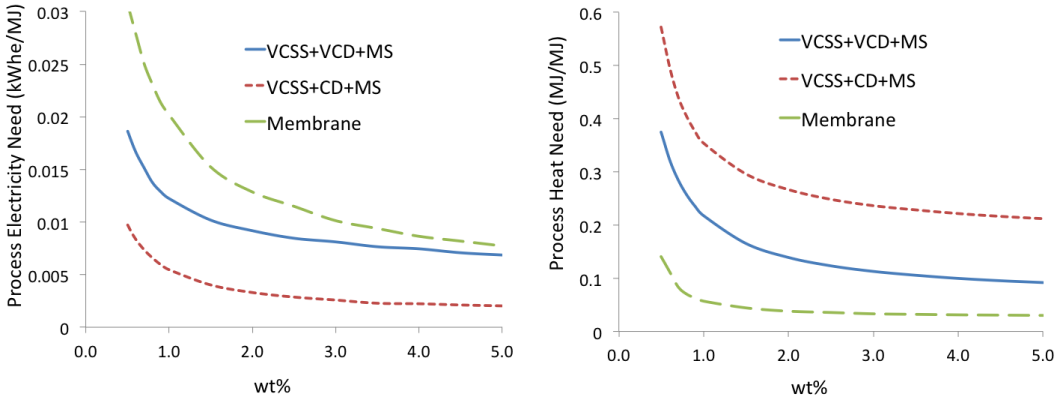


Figure 4: Electricity and heat need for ethanol separation processes

0.056 MJ/MJ_{EtOH}. This value is significantly lower than for benzene extractive distillation, but may be higher than that for membrane processes [155] [156].

Total Energy required for Ethanol Purification

The total electric energy requirement $E_e = 3.6 \times W_{adiabatic}$ in terms of MJ/MJ_{EtOH}, and total heat requirement $E_h = E_{evap} + E_{ms}$. Ethanol purification is the largest energy consumer in this cyanobacteria-to-fuel process. For our 1% example, the energy consumption in converting the initial 1% stream to fuel grade ethanol (99.7%) is 0.28 MJ/MJ with an 80% assumption for heat exchanger efficiency and about 0.19 MJ/MJ for a 90% efficiency. The 80% result is 15% lower than the model result for membrane-assisted vapor stripping (MAVS) of 0.33 MJ/MJ [155] [156]. With a 90% efficient heat exchanger, this VCSS-based process is significantly more efficient than the MAVS process. MAVS is a better choice at 5% initial concentration: our results are 0.13 and 0.11 MJ/MJ for 80% and 90%, respectively compared to 0.093 MJ/MJ for MAVS. For all concentrations in our range, both VCSS and MAVS are much better than conventional distillation [155] [156].

Figure 4 shows the total electricity and heat need for all three proposed processes under scenario 2 and assumes 80% heat exchange efficiency. For all ranges of initial concentration, VCSS plus membrane separation has the highest electricity requirement and lowest heat need; using VCD has both less electricity requirement and less heat requirement compared to conventional distillation.

2.3.3 Greenhouse Gas Emissions

Greenhouse gas emissions from electricity production depend on the fuel source and technology of generation. Full lifecycle greenhouse gas emissions have been estimated to be 1.03 kg CO_2e/kWh for coal and 570 g CO_2e/kWh for natural gas [77]. Two electricity sources are considered: grid electricity and on-site combined heat and power (CHP) using gas turbines. The planned Algenol-Dow pilot plant [138], planned for a location in Freeport Texas, will have grid electricity provided 37% from coal, 48% from natural gas, 12% from nuclear, 1.2% from wind, and 2% from other sources, implying a lifecycle greenhouse gas emission of about 700 g CO_2e/kWh. It will be shown later that use of this electricity would result in a greenhouse gas emission of 13.5 g CO_2e/MJ_{EtOH} for on-site electricity alone in the 1% case.

Natural gas has a greenhouse gas emission of 0.050 kg CO_2e/MJ [41]. If the natural gas boilers are 90% efficient, the greenhouse gas emissions per delivered MJ of natural gas are 55 g CO_2e/MJ. If all of the on-site process heat were provided by natural gas, the resulting on-site greenhouse gas emissions would be 11 g CO_2e/MJ_{EtOH} in the 1% example. Together with the 13.5 g CO_2e/MJ_{EtOH} from on-site electricity use and the 3.86 g CO_2e/MJ_{EtOH} from off-site emissions, the total for the grid-electricity system with natural gas heat is 28.3 g CO_2e/MJ_{EtOH} in the 1% example, easily meeting the renewable fuel requirement although not meeting the 20%-of-gasoline goal.

To further reduce the lifecycle energy use and greenhouse gas emissions, some of the process heat could be supplied by solar power. Solar power is intermittent, available for perhaps eight hours per day; existing industrial solar heat storage systems can extend this to up to 10 hours per day [74].

Gas turbines produce high-quality exhaust heat that can be used in CHP configurations to reach overall system efficiencies of 70 to 80 percent [49]. With an electricity generation efficiency of 38%, production of each kWh requires input energy of 9.5 MJ. With a steam

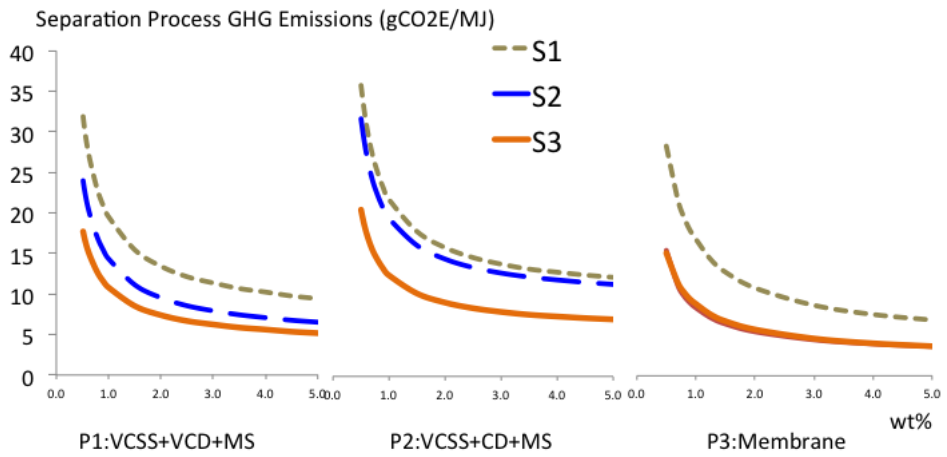


Figure 5: GHG emissions of different ethanol separation processes under three different energy supply scenarios

to power ratio 4 MJ/kWh, each kWh of electricity produced by natural gas will also generate 4 MJ of process heat [49]. The total energy use is estimated to be 0.36 MJ/MJ_{EtOH} if the initial ethanol concentration is 1%, and that the process electricity requirement and especially the process heat requirement decrease significantly as the initial ethanol concentration increases.

Figure 5 shows the GHG emissions of different ethanol separation processes compared under three different energy supply scenarios. The increased energy efficiency of a CHP system can provide significant greenhouse gas savings. It is obvious that using conventional distillation (P2) is the worst case scenario. There is little difference between S2 and S3 for membrane separation process (P3), because there is not enough heat required to utilize the potential of solar thermal in GHG emission reduction. Figure 5 shows that GHG emissions of ethanol separation process under S2 are always lower than S1.

2.4 Offsite, Pumping, and Water Use

In this section, life cycle stages other than the ethanol separation process are evaluated in terms of energy use and greenhouse gas emissions.

Mixing We assume a mixing system for the photobioreactors that is designed in much the same way as those for raceway ponds [136] with three primary functions: uniform suspension of the cyanobacteria, uniform distribution of nutrients, and efficient transfer of gases (CO_2 mainly) between the gas and liquid interface. The system is not designed to limit the effects of photo-saturation and/or photo-inhibition even though more rapid mixing could increase ethanol productivity, as the energy requirements for such a system are not practical for a fuel-production process [118] [62] [88]. Mixing within the photobioreactors is estimated to require 0.1 W/m_2 [136]. For the 1% case, this corresponds to 0.056 MJ/MJ_{EtOH}.

Oxygen Removal Removal of the oxygen will be accomplished through use of a compressor to extract the gas from the photobioreactor headspace, and a gas scrubber for oxygen removal, with an energy requirement of 0.0001 MJ/MJ_{EtOH}.

Water Consumption and Pumping We assume that the source water is pumped from a depth of 100 meters with 85% pumping efficiency and 95% motor efficiency, requiring 0.0066 MJ/MJ_{EtOH}; water pumping requirements will scale with this depth. Pumping of the water-ethanol chemical mixture into the separations system will require 0.004 MJ/MJ_{EtOH} under the same assumptions. Water sterilization, necessary for the initial fill of seawater and subsequent culture replacements, can be accomplished by ozonation with low energy requirements. Unlike growing algae in open ponds, we consider closed photobioreactors where water is not lost through evaporation. However, 3 moles of water are needed to produce 1 mole of ethanol, i.e. 0.926 liters of water for 1 liter of ethanol. This replacement water can be provided by reverse osmosis seawater desalination, which requires about 8 kWhe per 1000 gallon of water [1]. This is 9.5×10^{-5} kWh/MJ_{EtOH}

CO_2 **Source** Options for delivering CO_2 include gas delivery to the headspace of the photobioreactor, gas delivery to the culture, or CO_2-containing water delivered to the culture. Here we base our calculation on industrial CO_2 delivered to the headspace. Power

plant flue gas CO_2 could also be used, although additional clean-up of flue gas may be required. CO_2 transfer from the headspace to the culture is aided by the mixing system [136], the higher delivery concentration compared to atmospheric levels, and the higher sorption of CO_2 in seawater compared to fresh water.

Nitrogen and phosphorus fertilizers Emptying the photobioreactors once per year will create approximately 0.97 ton per hectare of waste biomass approximately 8% of which is nitrogen and 0.3% is phosphorus [45]. Accordingly, ethanol production of 56,000 l/hectare-yr corresponds to a nitrogen and phosphorus requirement of 0.065 g N/ MJ_{EtOH} and 0.0024 g P/MJ_{EtOH} respectively. If introduced all at once, the nitrogen use corresponds to three times the classic f/2 algal growth medium of Guillard [64] [53]. Because this is a no-harvest process, the cyanobacteria do not need to be continuously replenished, resulting in lower nutrient requirements than those for biofuel processes involving algal harvesting [35]. Greenhouse gas emissions associated with the nitrogen and phosphorus fertilizers are provided in the appendix.

Bioreactor Production The entire photobioreactor system is assumed to be replaced every five years. If the drained bioreactors are landfilled, the carbon in the plastic may be sequestered, with no net greenhouse gas emission from the bioreactors themselves. If the bioreactors are recycled, there will be avoided greenhouse gas emissions from the displacement of virgin polyethylene. We estimate 0.017 MJ/MJ_{EtOH} for the production energy and 1.0 g CO_2e/MJ_{EtOH} greenhouse gas emission contribution. Details are provided in the appendix A.

Waste Biomass Disposal The annual disposal of cyanobacteria is assumed to be managed by deep well injection, though other options are possible. Deep well injection will sequester the cyanobacterial biomass and may result in a net greenhouse gas reduction of approximately 2% of the system CO_2 emissions. Here we attribute zero sequestration and zero net emissions to cyanobacteria disposal.

Site preparation, ethanol distribution, and ethanol combustion These items make

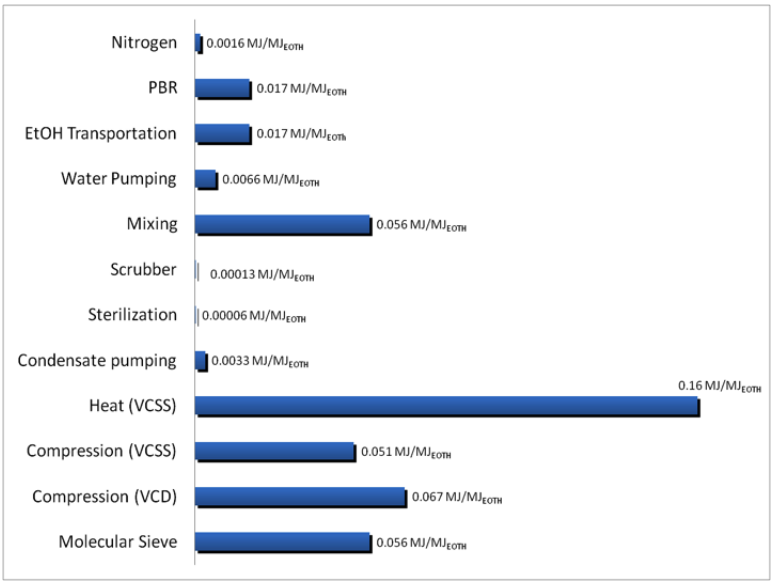


Figure 6: Baseline energy use for initial ethanol concentration of 1% assuming 80% heat exchange efficiency for the VCSS unit, for production of 56,000 l/ha/y

relatively small contributions and are discussed in SI.

Sensitivity to Ethanol Production Rate The ethanol production rate could be affected by photobioreactor geometry, cyanobacteria productivity, fertilizer requirements, mixing rate and mixing effectiveness, and other factors. The lifecycle elements that depend on the ethanol production rate include the bioreactor material, water, and fertilizer requirements, and the mixing and pumping energy. For the target production rate of 56,000 l/ha/y, a 50% lower or 50% higher production rate results in a total lifecycle energy use that is 9% higher or 5% lower, respectively.

2.5 Results and Discussion

Figure 6 shows that the ethanol separation process contributes most to the overall lifecycle energy use and GHG emissions. Therefore, it is important to carefully select separation process and energy supply scenarios to achieve the maximum potential in energy and GHG reduction.

The results and conclusions of this analysis are three fold:

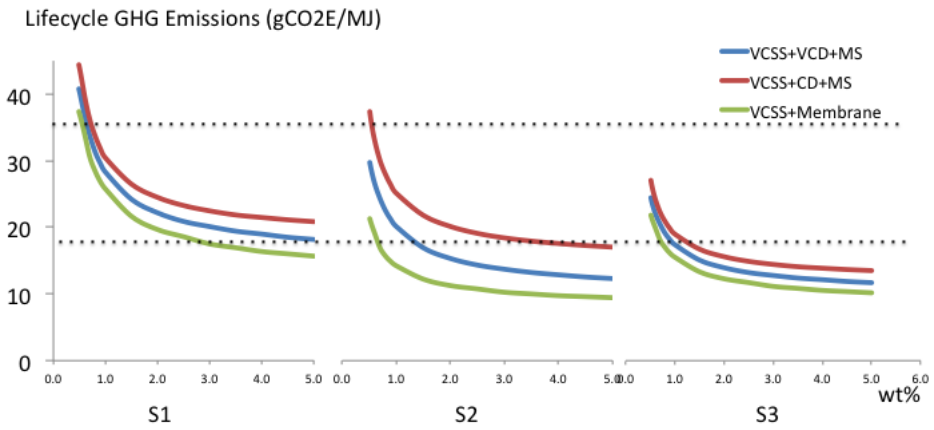


Figure 7: Lifecycle GHG emissions of different energy supply scenarios, based on an 80% heat exchange efficiency for the VCSS unit.

The first question to be answered is which ethanol separation process is the best. We compare them under different energy supply scenarios. As shown in figure 7, VCSS plus membrane separation (P3) is always better than the other two processes and using conventional distillation (P2) is the worst case. The difference is especially clear when combined-heat-and-power is used for energy supply. If grid electricity is used, then the difference of these three processes are not big if initial concentration is small, but the advantage of membrane separation increases as initial concentration gets bigger. When solar thermal is utilized, GHG emissions of all three candidate processes are reduced significantly, and the overall performances are similar.

The next question is which energy supply scenario is best and how much difference it makes. Figure 8 shows the total lifecycle greenhouse gas emissions for P1, P2, and P3 under three energy supply scenarios: grid electricity with natural gas for process heat; natural gas CHP; and natural gas CHP supplemented with solar heat. The figure shows that use of CHP provides a significant greenhouse gas emission benefit compared to grid-derived electricity, and that use of solar heat can provide significant additional reductions especially for P2. The use of solar heat is least beneficial for P3, because the heat requirement for P3 is too small which leads to less potential for reduction.

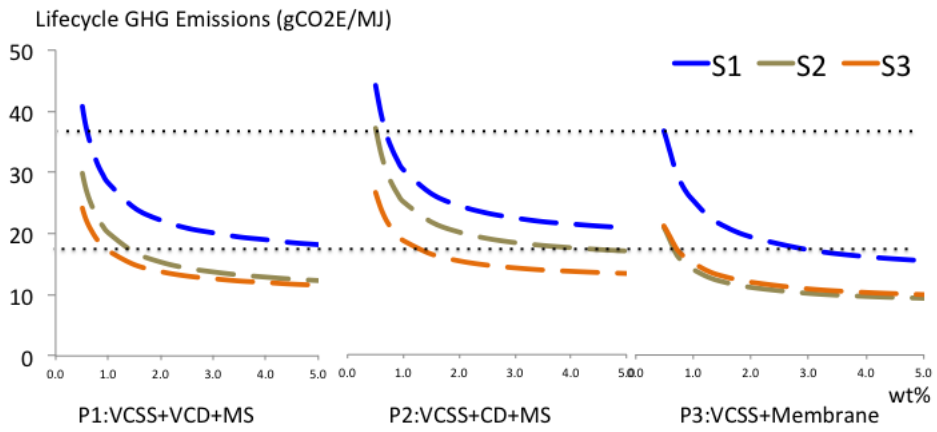


Figure 8: Lifecycle GHG emissions for P1, P2 and P3 under different energy supply scenarios, based on an 80% heat exchange efficiency for the VCSS unit.

And finally, how competitive is this ethanol production technology in terms of energy use and GHG emissions. The case of P1 is illustrated in Figures 9 as an example. Total greenhouse gas emissions from the algal ethanol process depend strongly on the initial ethanol concentration and on the energy system used to concentrate the ethanol. To meet the definition of a renewable biofuel under the U.S. Renewable Fuel Standard, the fuel must not exceed 36.5 g CO_2e/MJ_{EtOH}, which is 40% of the lifecycle greenhouse gas emission of gasoline. And to meet the DOE target goals of 20% of the gasoline emissions, the ethanol must have a lifecycle greenhouse gas emission of 18.3 g CO_2e/MJ_{EtOH} or less. Meeting the 40% requirement can be achieved under virtually all conditions and scenarios for P1. Meeting the 20% reduction target is more challenging. For an energy system consisting of moderately-low-carbon grid electricity and natural gas for process heat, the initial ethanol concentration needs to be at least about 4.0 to 4.5% dependent on the assumed level of heat exchange efficiency in the ethanol separation process. For an energy system based on natural gas CHP, the initial ethanol concentration needs to at least be about 1.0 to 1.2%. For the scenario involving solar heating, the initial concentration needs to be above 0.8% assuming 80% heat exchange efficiency and above 0.55% assuming 90% heat exchange efficiency.

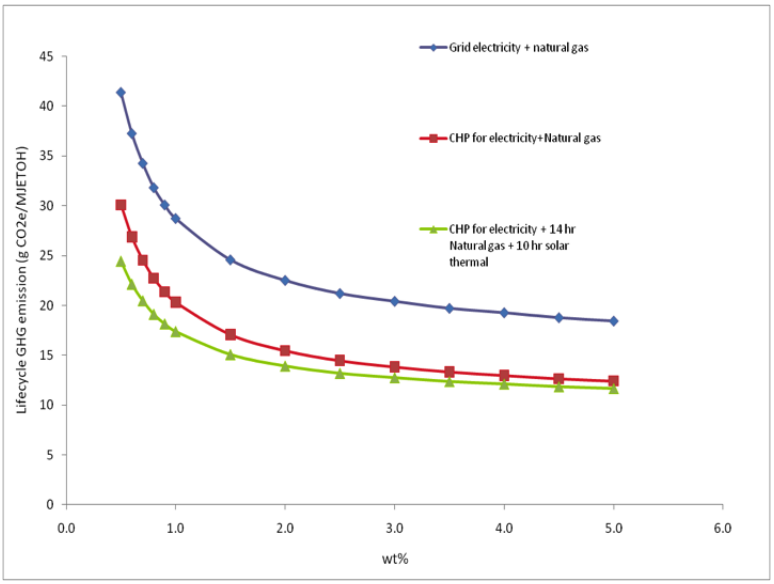


Figure 9: Lifecycle GHG emissions for P1 under under different energy supply scenarios, based on an 80% heat exchange efficiency for the VCSS unit.

The analysis is sensitive to the conservative assumption of 80% heat exchange efficiency in the vapor compression steam stripping and distillation process. More efficient heat exchange may be achieved in a practical system. A 90% assumption leads to net lifecycle energy consumption ranging from 0.38 down to 0.18 MJ/MJ_{EtOH} for the 0.5% to 5% initial concentration range for P1. This compares to 0.55 to 0.20 MJ/MJ_{EtOH} for the 80% assumption. Details are provided in the appendix A. Still higher heat exchange efficiencies are possible but would likely involve significantly higher capital investment.

An advantage of cyanobacteria-produced ethanol is the potential to locate production facilities on low-value, arid, non-agricultural land, and avoid competition with agriculture. Another advantage is the no-harvest strategy that has the potential for more energy efficient separations, lower fertilizer requirements, and lower water usage in comparison to other algal biofuel processes. We have shown that with sufficiently high initial concentrations of ethanol, fuel ethanol can be produced that has low net energy inputs and low lifecycle greenhouse gas emissions.

Chapter III

OPTIMAL PRODUCTION CAPACITY OF HIGHLY DISTRIBUTED BIOREACTOR-BASED FUEL SYSTEMS

3.1 Introduction

A chemical processing facility accepting raw material from a surrounding region may be shown to have an optimal size. One reason could be that the capital cost has positive economies of scale and raw material delivering cost increases superlinearly with the facility size. Biomass utilization facilities, such as corn ethanol and cellulosic ethanol plants, are of this nature. Biomass feedstocks are trucked from a wide geographic area to a central processing location, where the final product is produced. Because of the low density of biomass, biomass delivery is labor intensive and more expensive than the delivery of liquid and gaseous fuels such as petroleum or natural gas. It has been argued that there exists an optimal size for biorefineries based on corn and cellulosic materials because unit costs for processing decrease while feedstock transportation costs increases as the plant size increase [78] [169]. The optimal size of the plant depends on the competition between the scaling effects of these two factors. Studies have shown that the optimal capacity for corn and cellulosic ethanol occurs at a smaller scale, obtaining biomass from a distance as little as 15 miles [169]. Biorefineries converting lignocellulosic biomass through biochemical platform into ethanol were found to be optimally sized about 250 million gallon per year (MGY), and gasification-based biofuel plants were found to be in the range of 350 to 500 MGY [169]. They are larger than a grain ethanol plant which is estimated to have an optimal size around 60 to 80 MGY [169] [10] [56]. As a comparison, when raw material is abundant and unit transportation cost is constant, unit costs of the final product usually decrease monotonically with increasing plant size resulting in large optimal capacities.

Transportation fuels and chemical commodities derived from fossil fuels, for example, are generally produced from giant refineries. There are 148 operable refineries in the U.S., with operable capacity ranging from 31 million to 8.6 billion gallons per year [2].

These studies assume that all raw material, e.g. woody biomass, will be available at the same time for production. The decision models in these studies do not consider the timing of money spent and earned. This assumption is valid when full scale production can be achieved immediately with continuous raw material input, e.g. fossil fuel production where production is supported by mining from abundant reserves. The assumption is also reasonable when no benefit can be gained by starting the production earlier with smaller capacity. For example, in the case of ethanol production from terrestrial biomass such as switchgrass, if all contracted farmland starts growing switchgrass at the same time, the feedstock is available for harvesting or storage at the same time of the year. There is no reason to grow switchgrass on a smaller area, because no revenue can be generated earlier by doing so. Nevertheless, when the initial acquisition of the raw material flow is a step function of time, time value of money becomes an important factor in plant sizing.

This paper expands the plant sizing problem to consider situations where material flow, as input to the central production unit, has to be built up gradually over an extended time period and the production can only start when enough material flow is built. The time value of money can not be neglected in this case. Even though positive economies of scale in capital cost favors a larger facility size, it requires a longer time for production to start and therefore results in delayed revenue generation. This study is motivated by biofuel production systems based on algae cultivated in bioreactors [98] , where an initial chemical mixture is produced in individual bioreactors before being sent to a central processing unit. A large number of bioreactors, which may take several months to build, are needed to achieve a reasonable scale of production. This results in a tradeoff between earlier revenue generation by processing the chemical mixture earlier with a smaller production unit and taking advantage of economies of scale by building larger production units. Several

algae-based biofuel production systems and some pharmaceutical production systems are examples of this type of highly-distributed large-scale bioreactor system, which is defined formally in section 3.2.

This paper explores the optimal size of such a production system. Even though intensive liquid and gas transportation is needed because bioreactors are spread over a large geographic area, the sizing of the production unit is determined not only by the transportation cost of the chemical mixture and processing cost, but also by the time value of money. As a case study, this paper models and compares two different process designs of an ethanol production system from blue-green algae. An optimal production unit size that minimizes production cost is especially important for algae-based biofuel as an emerging technology. Therefore, production design that focuses on cost reduction and overall scale economy is an essential step towards commercialization.

3.2 Highly Distributed Bioreactor Based Production Systems

A *highly distributed bioreactor based production system* is a production system in which an intermediate chemical mixture is produced in continuous or batch mode in bioreactors that are distributed over a geographic area such that transport costs cannot be ignored. The chemical mixture is then collected, transported, and processed at one or multiple locations into final product with production rates at least an order of magnitude larger than those at the bioreactor level.

Examples include a range of types of algal biofuel and bio-product systems and could potentially also include bio-energy systems using land grown plants or non-biological solar-based systems. A bioreactor is the smallest unit in the system, where initial production is carried out. Each bioreactor requires inputs, such as nutrients and water, and generates outputs including the chemical mixture to be processed into final product. Because of the geographically highly distributed nature of this type of bioreactor system, both inputs and outputs of bioreactors usually require construction of a complex transportation network.

Therefore, transportation cost could be even higher than the value of the raw material. When the outlet flow rate of each bioreactor is fixed, an increase in the total production rate of the system requires increased number of bioreactors, and therefore larger geographical area and higher transportation costs.

Liquid and gas collection and distribution are important components of these systems and therefore are important parts of the system design process. Liquid and gas could be transported continuously on pipeline networks or in batches through trucks depending on the flow rate and costs. An important part of pipeline network design is the decision on diameter distribution at different levels, which is mostly determined by the tradeoff between capital cost (CAPEX) of the installed system and different operating costs (OPEX).

Chemical mixture collected from bioreactors may require multiple steps of processing to create the final product. Each processing step i increases the product concentration in the flow from α_{i-1} to α_i . The processing steps could be located in one central location or multiple locations. This depends on the cost structure and flexibility of each processing step. In some cases, a movable process column could be used to process flow at each bioreactor and then send the output to the next processing step. For example, a pyrolysis unit could be used to create pyrolysis oil at various locations and then the pyrolysis unit could be taken to a central location to process the oil into fuel. Let m_i be the number of units used in production step i and let $i = 1$ be the first process step whose inlet flow is direct flow from bioreactors and $i = n$ be the last process step whose product flow is the final product.The HDLB production system satisfies the following conditions:

1. **Product mass balance**: Let Q_i be the product flow of process step i and α_i be the concentration of final product in this product flow, then $m_{i-1}\alpha_{i-1}Q_{i-1} = m_i\alpha_i Q_i$

2. **Processing Cost**: Let c_i be the cost for one unit of process step i, the total processing cost of the system is $\sum_i m_i c_i$.

3. **Processing energy use and emissions**: Let e_i be the energy use of process step i

and g_i be the greenhouse gas emissions of one unit of process step i in the plant, the total processing energy use and processing emissions of the system is, respectively, $\sum_i m_i e_i$ and $\sum_i m_i g_i$.

4. **Flow conservation**: Let p_i, Q_i and f_i be, respectively, the inlet flow, the product flow and the co-product flow of process step i. The following statements are true:

   - Total inlet flow of process step i equals total product flow of the previous process step : $m_{i-1} Q_{i-1} = m_i p_i$
   - Inlet flow of process step i equals sum of product flow and co-product flow: $p_i = Q_i + f_i$
   - The net flow of all process steps is zero: $m_1 p_1 = m_n Q_n + \sum_i m_i f_i$

5. **Total land use**: Given the flow rate out of each bioreactor q and the land area occupied by one bioreactor $l \times w$, the total land area of the plant A is determined by the total production rate of the system $Q = m_n Q_n$ and initial product concentration at bioreactor levelα.

$$A = Nlw = \frac{m_n \alpha_n Q_n}{q} xy = \frac{\alpha_n lw}{q} \frac{Q}{\alpha} \tag{11}$$

where N is the total number of bioreactors in the system, and l and w are, respectively, the length and width of a bioreactor assuming it is rectangular.

3.3 Methodology

3.3.1 Problem Statement

The goal of this analysis is to provide a mathematical framework for designing the production capacity of a bioreactor based production system. Net present value of the project is used as the evaluation measure of different process designs. Discounted cash flow models are used to derive the net present value of the production system as a function of its

production rate and to show how different parameters influence this function and therefore influence the optimal production capacity.

Two types of production design are proposed as examples to demonstrate the method, as shown in figure 10. In one case, chemical mixture from each bioreactor is processed into the final product in an integrated process, which may include multiple process steps. These process steps may have the same economies of scale and capacity constraints, or the latter process steps may scale linearly. Here, we illustrate this case with an integrated two-step process, which include a local process A and local process B. The production capacity is determined by the inlet flow rate of process A, which determines the number of bioreactors serving one process A.

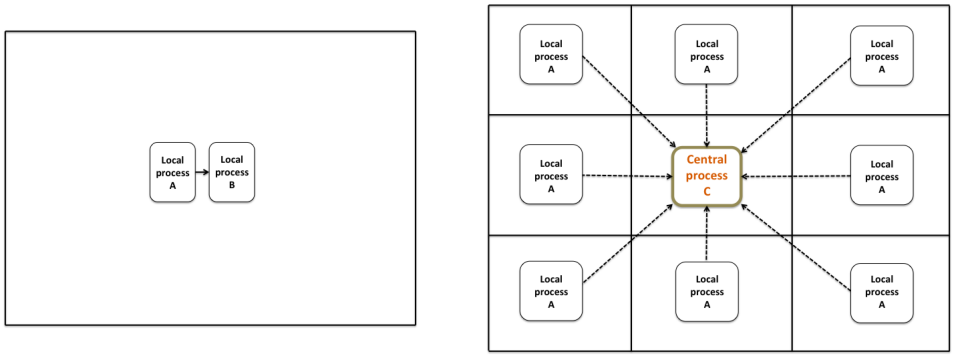


Figure 10: Layout of two production design scenarios, integrated processing, and distributed-centralized processing

In the second design scenario, a distributed-centralized production layout is allowed. Again, we illustrate this case using a two-step process that includes process A and process C. The chemical mixture can first be transported a short distance to local process A; and then the output of local process A is transported to a central location where it will be processed into the final product by process C. All local process A's supporting central process C are, for simplicity, assumed to be equivalently sized. The production, in this case, is determined by the inlet flow rate of local process A and the number of local process A's serving a central process C. Benefits can be gained by using the second design scenario

when, for example, local process A has a capacity limit that is much lower than central process C, so building a larger central process C can take more advantage of economies of scale. If the first process, e.g. local process A, can bring the material flow to a higher concentration, building the process in small capacity close to the bioreactors can reduce the flow needed to be transported to the central location and therefore reduce transportation costs. Here, local process A is set to be the same process for comparison of the two design scenarios, but it can be different processes for the two design scenarios depending on the particular processes required. Such distributed-centralized layouts for biofuel production have been modeled as mixed integer linear programming [176] [22] for cellulosic biofuels.

Comparison of the two designs can be based on production cost or profit generation. Therefore, models can be built for cost minimization or profit maximization policies. Here, we model the net present value of the plant as a function of the inlet flow rate of a local process A x, $y = f(x)$. The net present value of a project is to be maximized and taken as the only evaluation measure.

3.3.2 Model for Integrated Production Design

Notation and assumptions:

Let

- p be price per unit of the final product ;
- q be the flow rate out of each bioreactor in terms of volume per time period;
- n be the number of bioreactors that can be built per time period;
- α be the initial concentration of product element in the chemical stream out of each bioreactor;
- α_1 be the concentration of product element in the output stream of local process A;
- b be the capital cost of each bioreactor;

- δ be the operational cost of each bioreactor per time period;
- $C^A = C_0^A(\frac{x}{Q_0^A})^{s_A}$ be the capital cost of building a local process A, where C_0^A is the benchmark CAPEX with inlet flow rate Q_0^A and s_A is its scaling exponent.
- $C^B = C_0^B(\frac{\frac{\alpha}{\alpha_1}x}{Q_0^B})^{s_B}$ be the capital cost of building a local process B, where C_0^B is the benchmark CAPEX with inlet flow rate Q_0^B .
- $C^C = C_0^C(\frac{\frac{\alpha}{\alpha_1}x}{Q_0^C})^{s_C}$ be the capital cost of building a central process C, where C_0^C is the benchmark CAPEX with inlet flow rate Q_C .
- $S = S_0(\frac{x}{Q_0^S})^s$ be the capital cost of the storage tank, where S_0 is the storage tank CAPEX for a local process with inlet flow rate Q_0^S;
- γ be ratio of minimum flow to maximum flow which a production equipment can handle, i.e. the inverse of the turndown ratio ;
- $C^P = C_0^P(\frac{x}{Q_0^P})^z$ be the CAPEX of pipeline, where C_0^P is the CAPEX of pipeline for a local process with inlet flow rate Q_0^P and z is the pipeline scaling exponent;
- C^I be the sunk costs to be paid at time zero per unit of the final product produced;
- o be the operating cost that scales linearly with production rate;
- $e = e_0\alpha^{-\epsilon}$ be the unit energy cost that scales with initial concentration;
- T be the project life time once operation starts;
- $d = 1 + i$ be the effective discount rate in one time period, where i is the monthly compounded interest rate.

and let

- t_0 be the time when the production equipment is built and starts operation at minimal rate;

- t_s be the time at which flow rate from the bioreactors alone reaches the minimum required inlet flow rate of the production equipment;

- t_f be the time when the production equipmen first reach full capacity;

What factors to include in the plant decision model depend on the nature of the considered projects and the planner's perspective. Constraints such as technical feasibility and the planner's objective judgement can be added to the model. The production cost components included in this analysis and assumptions of their scale economies are as follows:

Capital and operational cost of bioreactors

For highly-distributed large-scale bioreactor systems, a large number of bioreactors have to be built to support the production. The costs associated with each bioreactor typically include bioreactor material cost, labor cost in building the bioreactor, and energy cost in running the bioreactor. Here, we assume that there is a fixed capital cost for each bioreactor built and the operational cost per time period of each bioreactor is a constant. Before the completion of all bioreactors, a fixed capital cost is spent at the beginning of each time period, which is the sum of the capital cost of all bioreactors to be built in that time period; an operational cost is spent at the end of each time period, which is the sum of operational cost per period for all the bioreactors built until that time.

The speed at which the bioreactors are built could significantly affect the timing of production. Obviously, more labor can be put into building the bioreactor to speed up the process. However, this approach may not be taken by the planner because of the inability to learn by previous mistakes. Hence, the planner will always decide on a reasonable number of bioreactors to be built per time period based on the tradeoff between the learning effect and the cost of labor. The building speed may increase as experience increases. Nevertheless, this decision is out of the scope of this paper, and the number of bioreactors that can be built per time period is taken as a exogenous parameter in the analysis.

The total cost of building the bioreactors at time h includes capital cost of bioreactors

for the next period and the operational cost for the past period. $\bar{B}_h = b + \delta h$, from time zero to time t_f. The net present value of total capital and operational cost of the bioreactors is

$$\hat{B}_h = \sum_{h=0}^{t_f} \frac{\bar{B}_h}{d^h} = \sum_{h=0}^{t_f} \frac{b + \delta h}{d^h} \tag{12}$$

Capital cost of production equipment

The process units are usually the biggest contributors to the capital and operational cost of a production system. In the literature, the capital cost of many processes are typically assumed to be dependent on its inlet flow rate:

$$C = C_0(\frac{Q}{Q_0})^s \tag{13}$$

where C_0 is the cost of the process for a benchmark flow rate Q_0, and s is the dimensionless scale factor that reflects the economies of scale in production, which theoretically can take any value between 0 and 1 and usually ranges from 0.6 to 0.9 for biomass electricity plants. Though not many cellulosic ethanol plants are in commercial operation, studies on corn ethanol production suggest a similar range of scale factor values. Gallagher et al. [56] have estimated s to be 0.836 for the dry mill ethanol industry based on a USDA cost of production survey. In some cases, s can also be treated as a function of the capacity Q [78]. Even though the unit capital cost decreases as capacity increases in this formula, which suggests a large optimal capacity, many researchers have shown that the decrease is quite small beyond some small capacity. As the scale factor approaches to 1, the cost becomes less sensitive to the inlet flow rate. When multiple steps of separation are required, each separation step might have different scaling exponents.

Another important factor in the cost structure of biomass plants is the operational cost, which may include energy costs, labor costs, and raw material purchasing and handling cost. For example, gasification is believed to require more capital cost but less energy input per unit of biomass processed compared with direct combustion [27] for biomass power plants [169]. Most researchers have assumed the unit operating costs to be constant, but

they can also scale with plant size [131]. In this analysis, these costs are assumed to be constant for each unit of the final product produced. Therefore, it is not directly expressed in the model, but instead, subtracted from the price of each unit of the final product.

A production unit usually requires its inlet flow rate to be above a certain threshold to start operation. Hence, it is important to consider the time at which operation starts and its effect on the profits of the plant. If the equipment stands idle for an extended period to wait for bioreactors to be built up, revenue is lost [114]. Therefore, this analysis assumes that the capital cost of a production unit is only spent when enough inlet flow rate has been built up. Processes with large capacity require more bioreactors and infrastructure to support the inlet flow rate and hence requires longer time until operation.

Energy cost

Energy cost is usually assumed to be scale linearly with production capacity in the literature for economic analysis. The energy cost of distillation, however, has been shown to scale not only with production rate, but also the composition of chemical mixture in the inlet flow [155]. It has been shown that, for ethanol water separation, the energy cost decreases as the ethanol concentration in the inlet flow increases. Therefore, we assume that energy cost per unit of final product e is given as follows:

$$e = e_0 \alpha^{-\epsilon} \tag{14}$$

where e_0 is a constant, and $-\epsilon$ is the scaling exponent of the energy cost.

Capital cost of storage tanks

A steady state flow of chemical mixture is expected to be generated shortly after a bioreactor is built. A storage tank is needed to store the chemical mixture before enough bioreactors have been built to guarantee a continuous inlet flow to start the production unit. The storage tank should be built at the beginning of the project, and its size is dependent on the size of the inlet flow rate of local process A.

Capital and operational cost of piping

Biomass feedstocks for corn and cellulosic biofuel plants are usually assumed to be transported by trucks, though pipeline transportation of biomass as sludge has also been studied [86]. In the biomass energy literature, the cost of delivering biomass from the surrounding farmland is believed to increase with production rate due to the low density of biomass and the scarcity of resources. As a result, the optimal plant size is also dependent on the biomass yield in the region [27]. The delivery cost can be estimated based on an assumed geometry of the biomass supply region surrounding the facility [78] [110] [57]. The delivery cost C^P has been assumed to follow equation 15

$$C^P = C_0^P(\frac{Q}{Q_0^P})^z \tag{15}$$

where C_0^P is the delivery cost for benchmark production rate Q_0^P and z is the scaling exponent.

and A detailed cost profile for harvest, purchase, and transport of biomass is provided in [87] and [27].

For production systems based on bioreactors, the chemical mixture from each bioreactor could be transported to a local process A continuously through pipeline network or in batches by trucks. Trucking can be more cost effective only when the flow rate of each bioreactor is very slow, but pipeline is the most commonly used method for liquid and gas transportation. The scaling exponent z is estimated to be 1.5 to 2 for truck transportation [169] [109] and in the appendix B, we estimate that z is 1.1 for pipeline transport. Here, we assume that pipeline is used for transportation, which incurs a capital cost for the installation system, which follows equation 15.

Initial investment

Initial investment when building a new plant may include land cost, administration cost, laboratory cost, labor cost, etc. We assume that the sunk cost at the beginning of the project increases linearly with the production capacity of the plant. So, the total sunk cost at time zero is denoted by $I = C^I Q = C^I \alpha x$.

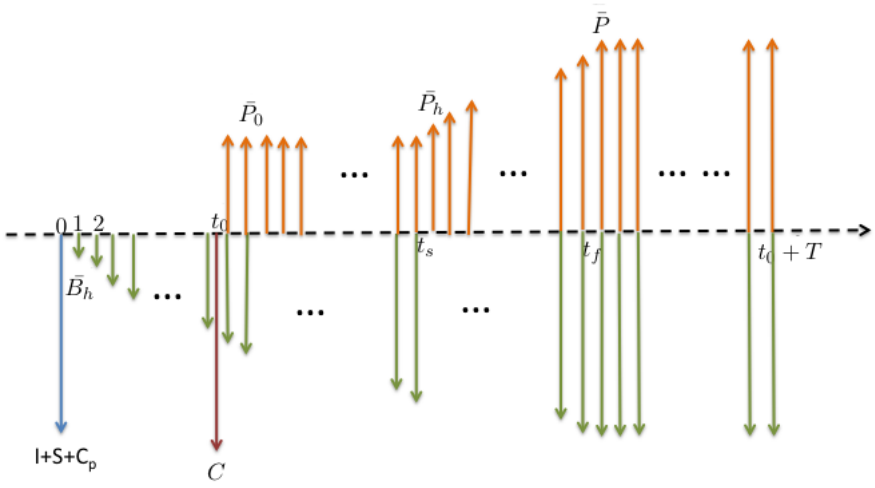


Figure 11: Cash flow over project life time for integrated processes

Revenue generation

Revenue is gained through sales of the final product. We assumes constant revenue per unit of final product produced and infinite demand. In the first design scenario shown in figure 10, all processing of the chemical mixture into final product is completed at one location. A two-step production is required to process chemical mixtures into final product. The stages of the intermediate chemical mixture are characterized by the concentration of the main product element in the mixture, e.g. ethanol in ethanol-water solution in the case study. The initial concentration α from the bioreactor is raised first by local process A to α_1 and then by local process B to one. Let x be the inlet flow rate of local process A, the production capacity is then $Q = \alpha x$. The net present value of the system is, therefore, $y = f(x) = f(\frac{Q}{\alpha})$.

Figure 11 shows the cash flow and timing of the integrated design.

- At time zero, the plant invests in a sunk cost I. In the mean time, S is spent to build an intermediate tank to store the chemical mixture generated before the production units start operation; and C^P is spent to build the pipeline installation system. At the same time, the plant starts building bioreactors.

- The capital and operational cost of bioreactors are invested in batches on a periodic basis. In period h, a total cost of $\bar{B}_h$ is spent. Bioreactors are built at a speed of n.

- At time t_0, the local process A and local process B are built and the capital costs are spent. The local process A starts operation at minimum flow rate γx with existing bioreactors and the mixture stored in the intermediate tank. The inlet flow rate of local process A will stay at γx until time t_s and revenue are collected at the end of each time period.

- At time t_s, the previously stored mixture is depleted but the bioreactors built already reaches γx by themselves. Revenue will start increasing as processing rate increases with increasing bioreactors until the local process A reaches full capacity;

- At time t_f, local process A reaches full capacity. Revenue $\bar{P}$ will be generated on a periodic basis;

- At time $t_0 + T$, the basic production system reaches the end of life.

Determine the time to build local process A as a function of is capacity

The local processes A and B should start operation as soon as possible but not until the inlet flow rate can be guaranteed at a minimal of γx without interruption. It is not hard to prove that in order to start the local process as soon as possible, the previously stored mixture will be processed in a reversed fashion as it was accumulated. Once the stored mixture is depleted, the number of bioreactors should be exactly enough to support a flow rate of γx and not more. Because, otherwise the local process can be started earlier.

As shown in figure 12, the flow rate of chemical mixture in the system increases with time as more bioreactors are built up over time. Line CA represents the level γx at which the local process can start operation. Line OA represents the flow rate from the bioreactors as a function of time. The slope of OA, q, is the speed at which flow rate is built up. Triangle OED defines the amount of chemical mixture generated from time zero to t_0, i.e.

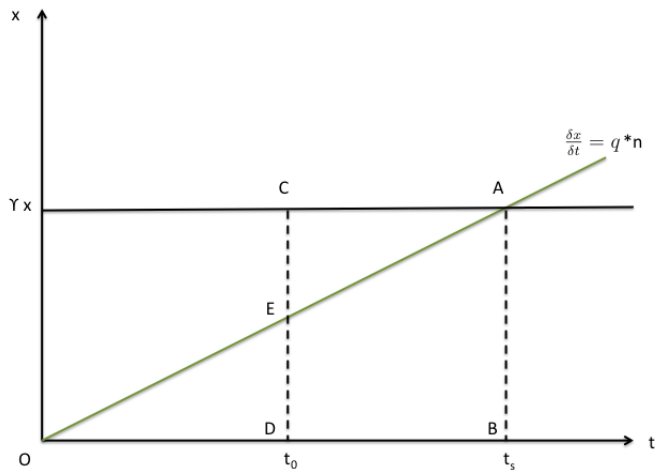


Figure 12: Flow rate of α chemical mixture in the system as a function of time

the volume in the storage tank at time t_0. DEAB represents the amount of chemical mixture generated from the bioreactors from time t_0 to t_s. If the process starts operation at time t_0 to process chemical mixture at a rate of γx, then square CABD represents the amount of mixture processed by the local process from t_0 to t_s. The amount of mixture ever generated by t_s has to be at east as much as the mixture processed by t_s, i.e. the area of triangle OAB is greater than or equal to the area of square CABD.

$$\frac{1}{2} t_s \gamma x \geq (t_s - t_0)\gamma x \tag{16}$$

According to equation 16, $t_0 \leq \frac{1}{2} t_s$. Since we want to start the production process as soon as possible, $t_0 = \frac{1}{2} t_s$

From graph 12, $t_0 = \frac{1}{2} t_s$ indicates that the triangle OED and triangle CEA are equivalent, which means CE equals ED. The following relations are true:

$$nq = \frac{\delta x}{\delta t} = \frac{\gamma x}{t_s} = \frac{\frac{1}{2}\gamma x}{t_0} \tag{17}$$

From equation 17, we get t_s and t_0 as functions of x,

$$t_s = 2t_0 = \frac{\gamma x}{nq} \tag{18}$$

And the time until the local process reaches full capacity is

$$t_f = \frac{x}{nq} \tag{19}$$

From time zero to time t_0, the flow generated from bioreactors is stored in the storage tanks. As shown in figure 12, the capacity of the storage tank, Q_x^S, should equal to the total amount of flow needed to be stored is the total area of triangle OED:

$$Q_x^S = \frac{\gamma x t_0}{4} = \frac{\gamma^2 x^2}{8nq} \tag{20}$$

From t_0 to t_s, revenue is generated at a rate of $\bar{P}_0 = \tilde{p}\alpha\gamma x$, where $\tilde{p} = p - o - e$. The NPV of revenue generated in this period is,

$$\hat{P}_0 = \sum_{h=t_0}^{t_s} \frac{\tilde{p}\alpha\gamma x}{d^h} \tag{21}$$

From t_s to t_f, the processing rate keeps increasing at a speed of nq per period starting from γx. At time h , the revenue generated is given by,

$$\bar{P}_h = \tilde{p}\alpha[\gamma x + (h - t_s)q] \qquad h = t_s + 1, t_s + 2, ..., t_f \tag{22}$$

The NPV of the revenue generated from t_s to t_f can be calculated by,

$$\hat{P}_h = \sum_{h=t_s+1}^{t_f} \frac{\bar{P}_h}{d^h} = \sum_{h=t_s+1}^{t_f} \tilde{p}\alpha \frac{\gamma x + (h - t_s)q}{d^h} \tag{23}$$

Once the process reaches full capacity at time t_f, the revenue generated at the end of each time period is $\bar{P} = \tilde{p}\alpha x$.

The net present value of the revenue from the start of full scale production until the end of the project life can be calculated as

$$\hat{P} = \sum_{h=t_f+1}^{t_0+T} \frac{\tilde{p}\alpha x}{d^h} \tag{24}$$

The net present value of the project can be calculated as

$$\begin{aligned} y_1 = f_1(x, \mathbf{p_1}) = & \sum_{h=t_0}^{t_f} (p - o - e_0\alpha^{-\epsilon}) \frac{\alpha\gamma x}{d^h} + \sum_{h=t_s+1}^{t_f} (p - o - e_0\alpha^{-\epsilon}) \frac{\alpha(h - t_s)nq}{d^h} \\ & + \sum_{h=t_f+1}^{t_0+T} (p - o - e_0\alpha^{-\epsilon}) \frac{\alpha x}{d^h} - S_0 (\frac{x}{Q_0^S})^s - C^I \alpha x - \sum_{h=0}^{t_f} \frac{n(b + \delta h)}{d^h} \\ & - \sum_{h=t_f+1}^{T} \frac{n\delta t_f}{d^h} - C_0^P (\frac{x}{Q_0^P})^z - \frac{C_0^A (\frac{x}{Q_0^A})^s}{d^{t_0}} - \frac{C_0^B (\frac{\frac{\alpha}{\alpha_1} x}{Q_0^B})}{d^{t_0}} \end{aligned} \tag{25}$$

$y_1 = f_1(x, \mathbf{p_1})$ can be derived from equation 25, equation 18 and equation 19, where $\mathbf{p_1}$ is a vector of parameters:

$\mathbf{p_1} = (p, q, n, \alpha, b, \delta, C_0^A, C_0^B, S_0, \gamma, s, C_0^P, z, C^I, o, e_0, \epsilon, T, d)$.

With known $\mathbf{p_1}$, $x_1^* = \arg\max f_1(x, \mathbf{p_1})$ can be derived by Karush-Kuhn-Tucker (KKT) conditions. However, some of the parameters could be uncertain or stochastic. In section 3.3.3, global sensitivity analysis is used to enable the decision maker to understand the degree of confidence in the decision and to identify the most influential system parameters.

3.3.3 Model for Distributed-Centralized Production Design

For design scenario two, where a central process C is supported by multiple local process A's, we model the system with two decision variables: x is the capacity of a local process A; and k is the number of local process A's supporting one central process C. The inlet flow rate of the central process C, which is the sum of the outlet flow rate of all the local process A's, is then $x_c = kx\frac{\alpha}{\alpha_1}$.

Here, we introduce three additional parameters:

- $C^{P1} = C_0^{P1}(\frac{x}{Q_0^{P1}})^z$ is the CAPEX of pipeline that connects local process A's to the central process C;

- F_A is the fixed cost term for a local process A unit;

- t_v is the time it takes to build on local process A unit;

Notice that the fixed cost term was neglected for the integrated production design as well as for the central process C. This is because when the production capacity is very large, the fixed cost is very small compare to the overall cost of the production equipment. In distributed-centralized production design, however, a fixed cost term is necessary to prevent from a solution with very small local process A columns.

If we do not consider the time value of money, the optimal size of a local process A should be determined by the tradeoff between transportation cost and economies of scale in processing. Nevertheless, in building the model for design scenario one, we have shown that the time value of money plays significant role in sizing of the local process A. If the output of the local process A in the second design scenario has an immediate revenue, then the optimal size of a local process A should be determined in the same way as design scenario one. However, when a local process A is connected with a central process C, revenue is only generated when central process C starts operation. For simplicity, we assume that a local process A starts operation only when its full capacity is reached. Assume that the local process A's are built sequentially, then the inlet flow rate of the central process increases in steps as more local process A's start operation.

In this case, t_s, t_0, t_v, t_f can be all be represented as functions of x and k:

From equation 17, we get t_s and t_0 as functions of the inlet flow rate,

$$t_s = 2t_0 = \frac{\gamma k x}{nq} \tag{26}$$

And the time until the central process reaches full capacity is

$$t_f = \frac{kx}{nq} \tag{27}$$

Assuming local process A's are built sequentially, the time it takes to build one local process A is determined by the inlet flow rate of local process and and the bioreactor building speed.

$$t_v = \frac{x}{nq} \tag{28}$$

The net present value of the project can be calculated as

$$\begin{aligned} y_2 = f_2(k, x, \mathbf{p_2}) = & \sum_{h=t_0}^{t_f} (p - o - e_0\alpha^{-\epsilon}) \frac{\gamma k x \alpha}{d^h} + \sum_{j=1}^{v} \sum_{h=t_s+(j-1)t_v+1}^{t_s+jt_v} (p - o - e_0\alpha^{-\epsilon}) \frac{\alpha x}{d^h} j \\ & \sum_{h=t_f+1}^{t_0+T} \frac{(p - o - e_0\alpha^{-\epsilon}) k \alpha x}{d^h} - S_0 (\frac{kx}{Q_0^S})^s - C^I k \alpha x - \sum_{h=0}^{t_f} \frac{n(b + \delta h)}{d^h} \\ & - \sum_{h=t_f+1}^{T} \frac{n\delta t_f}{d^h} - kC_0^{P1} (\frac{x}{Q_0^{P1}})^z - \frac{C_0^C (\frac{kx\frac{\alpha}{\alpha_1}}{Q_0^C})^s}{d^{t_0}} - \sum_{j=1}^{k} \frac{F^A + C_0^A (\frac{x}{Q_0^A})^s + C_0^P (\frac{x}{Q_0^P})^z}{d^{jt_v}} \end{aligned} \tag{29}$$

$y_2 = f_2(k, x, \mathbf{p_2})$ can be derived from equation 29, equation 26 and equation 27, where $\mathbf{p_2}$ is a vector of parameters:

$\mathbf{p_2} = (p, q, n, \alpha, b, \delta, F^A, C_0^A, C_0^C, S_0, \gamma, s, C_0^P, C_0^{P1}, z, C^I, o, e_0, \epsilon, T, d)$.

3.3.4 Global sensitivity analysis

The analysis of the sensitivity of the model output to variation of the input parameters is of significant importance to the evaluation of model performance, especially for the multi-parametric models proposed in this analysis. Complex models often have the problem of over-parameterization. Therefore, sensitivity analysis methods that aim to reduce the number of parameters are commonly used.

Local sensitivity analysis investigates the impact of parameters on model output based on changes in parameter values only very close to the nominal values. Local sensitivity

methods can be used when input parameters are known with little uncertainty, and partial derivative of the output function with respect to input parameters can be computed numerically by performing multiple simulations varying input parameters around a nominal value. However, when input parameters are very uncertain, local sensitivity analysis is not sufficient. Comparatively, for a generic model such as $y = f(\mathbf{p})$, where $\mathbf{p} = (p_1, p_2, ..., p_k)$ is the set of input parameters, global sensitivity analysis evaluates the relevance of x_i on Y when the entire distributions of $\mathbf{p}$ and output Y are considered [20]. The global sensitivity method used in this paper is based on Sobol's indices, a measure of importance of the fractional contribution of the input parameters to the variance of the model output. Sobol's method, along with the Fourier amplitude sensitivity test (FAST) and their extended versions, are global sensitivity measures based on decomposition of model output variance.

Models $y_1 = f_1(x, \mathbf{p_1})$ and $y_2 = f_2(k, x, \mathbf{p_2})$ are highly nonlinear and non-monotonic, and the distribution functions of the input parameters range over many orders of magnitude, which makes sensitivity analysis especially difficult [129]. An analytical form of optimal capacity as a function of all the parameters does not exist. Therefore, the Monte Carlo method is used to generate parameter estimates from an assumed distribution, and the model is executed once for each set of Monte Carlo estimates. For such nonlinear non-monotonic models, the most appropriate methods are based on decomposition of model output variance, such as the Sobol method.

We apply an extension of Sobol's method using the Monte Carlo procedure proposed by Satelli [128]. We calculate a first order Sobol's index and the corresponding total effect for each parameter as well as some higher order Sobol's indices to identify the non-influent parameter set. Total effect indices account for all the possible synergetic terms between the given parameters and all the others. The difference between the first order indices and the total effect is a measure of the nonlinearity of the model. The nonlinearity of a model also depends on which output variable is considered.

First, two input sample matrices are generated through Monte Carlo simulation:

$$\mathbf{M}_1 = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,k} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,k} \end{pmatrix}, \mathbf{M}_2 = \begin{pmatrix} p'_{1,1} & p'_{1,2} & \cdots & p'_{1,k} \\ p'_{2,1} & p'_{2,2} & \cdots & p'_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ p'_{n,1} & p'_{n,2} & \cdots & p'_{n,k} \end{pmatrix}$$

where n is the sample size used for the Monte Carlo estimates.

For each row of M_1 or M_2, the model $y = \mathbf{p}$ can be executed once to get a particular y. Therefore, by executing the model n times on sample M_1, the expected value of y, $E(y)$, can be computed from all the y's generated. This can also be done on sample M_2.

In order to estimate the sensitivity measure for a generic parameter p_j, i.e.

$$S_j = \frac{V(E(y \mid p_j))}{V(y)} = \frac{U_j - E^2(y)}{V(y)} \tag{30}$$

$$U_j = \int E^2(y \mid p_j = \tilde{p}_j) p_j(\tilde{p}_j)\, \mathrm{d}\tilde{p}_j \tag{31}$$

U_j can be obtained from values of y computed on matrices M_1 and N_j , where N_j is defined as:

$$\mathbf{N}_j = \begin{pmatrix} p'_{11} & p'_{12} & \cdots & p_{1j} & \cdots & p'_{1k} \\ p'_{21} & p'_{22} & \cdots & p_{2j} & \cdots & p'_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ p'_{n1} & p'_{n2} & \cdots & p_{nj} & \cdots & p'_{nk} \end{pmatrix}$$

i.e. by

$$\hat{U}_j = \frac{1}{n-1} \sum_{r=1}^{n} f(p_{r1}, p_{r2}, ..., p_{rk}) f(p'_{r1}, p'_{r2}, ..., p'_{r(j-1)}, p_{rj}, p'_{r(j+1)}, ..., p'_{rk}) \tag{32}$$

Take matrix M_1 as the sample matrix, and M_2 as the re-sample matrix, then U_j is obtained from products of values of f computed from the sample matrix times values of f computed from N_j , i.e. a matrix where all factors except x_j are re-sampled. In this way the

computational cost associated with a full set of first order indices S_i is $n(k+1)$. One set of n evaluations of f is needed to compute $E(y)$, and k sets of n evaluations of f are needed for the second term in the product.

We adopt Homma and Satalli's method in estimating $E(y)$ using both matrix M_1 and M_2 [70],

$$\hat{E}^2 = \frac{1}{n}\sum_{r=1}^{n} f(p_{r1}, p_{r2}, ..., p_{rk}) f(p_{r1}^{'}, p_{r2}^{'}, ..., p_{rk}^{'}) \tag{33}$$

In order to identify a subset of the k factors that could account for most of the variance of the output y, the parameters can be partitioned into a target set $\mathbf{u} = (p_{i_1}, p_{i_2}, ..., p_{i_m})$ and the remaining set $\mathbf{v} = (p_{l_1}, p_{l_2}, ..., p_{l_{k-m}})$

$$S_{\mathbf{u}} = \frac{V(E(y \mid \mathbf{u}))}{V(y)} = \frac{U_{\mathbf{u}} - E^2(y)}{V(y)} \tag{34}$$

$$S_{\mathbf{u}}^{tot} = \frac{V(E(y \mid \mathbf{u})) + V(E(y \mid \mathbf{u}, \mathbf{v}))}{V(y)} = \frac{V(y) - V(E(y \mid \mathbf{v}))}{V(y)} \tag{35}$$

where $S_{\mathbf{u}}$ is the first order Sobol's indice of set $\mathbf{u}$, $V(E(y \mid \mathbf{u}))$ is the first order effect (i.e. variance corresponding to set $\mathbf{u}$), $S_{\mathbf{u}}^{tot}$ is the total effect indice, $V(E(y \mid \mathbf{u}, \mathbf{v}))$ is the variance corresponding to the interaction between set $\mathbf{u}$ and set $\mathbf{v}$. If $V(y) \cong V(y \mid \mathbf{u})$ (i.e. $S_{\mathbf{v}}^{tot} << 1$), then set $\mathbf{v}$ is non-influent. In this case, $f(\mathbf{p})$ mainly depends on set $\mathbf{u}$, and all parameters in set $\mathbf{v}$ can be fixed in subsequent analysis [128] [143] .

To obtain $S_{\mathbf{u}}$ and $S_{\mathbf{u}}^{tot}$, we only have to calculate $V(y)$, $U_{\mathbf{u}}$, $U_{\mathbf{v}}$. The Monte Carlo estimates of $U_{\mathbf{v}}$ can be calculated by,

$$\hat{U_{\mathbf{v}}} = \frac{1}{n-1}\sum_{r=1}^{n} f(p_{ri_1}, p_{ri_2}, ..., p_{ri_m}, p_{rl_1}, p_{rl_2}, ..., p_{rl_{k-m}}) f(p_{ri_1}^{'}, p_{ri_2}^{'}, ..., p_{ri_m}^{'}, p_{rl_1}, p_{rl_2}, ..., p_{rl_{k-m}}) \tag{36}$$

It is easy to see that 30 is a special case of 34 where one set contains only one parameter and the other contains the rest. Therefore, to calculate the total effect of a single parameter

S_j^{tot}, one can simply take $\mathbf{u} = (p_j)$ and $\mathbf{v} = (p_1, p_2, ..., p(j-1), p(j+1), ..., p_k)$ and use equation 35.

3.3.5 Conditional-Value-at-Risk Optimization

Like most engineering problems, the model developed for this analysis involves uncertain parameters. An optimal decision is decided based on evaluation of the random variables. Different criteria can be used for the selection. In this paper, we consider using Conditional-Value-at-Risk as a risk measure for optimal decision making.

Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR) are methodologies developed by the financial industry to provide quantification for a company's portfolio's exposure to risk. VaR is defined as a threshold value such that the probability that the loss exceeds this value is the given probability level. CVaR is defined as the conditional expectation of losses given that the loss exceeds a threshold value (VaR). The limitations of VaR, especially its lack of subadditivity and convexity, have been widely recognized in the literature [7] [8]. CVaR, as an alternative risk measure, provides coherent properties such as convexity, positively homogeneous [115]. Moreover, minimizing CVaR typically leads to a portfolio with a small VaR.

Let $\mathbf{x}$ be the decision variable, and $g(\mathbf{x}, \mathbf{p})$ be a loss function having a distribution in IR induced by that of $\mathbf{p}$. In this paper, we consider the loss function to be the net loss of the profit defined by the negative of the net present value. Therefore, for an integrated design, the loss function is $g(\mathbf{x}, \mathbf{p}) = -f_1(\mathbf{x}, \mathbf{p_1})$.The underlying distribution of $\mathbf{p}$ is denoted by $g(\mathbf{p})$. The probability of $g(\mathbf{x}, \mathbf{p})$ not exceeding a threshold of θ is given by

$$\psi(\mathbf{x}, \theta) = \int_{g(\mathbf{x},\mathbf{p})\leq\theta} p(\mathbf{p}) d\mathbf{p} \tag{37}$$

In general $\psi(\mathbf{x}, \theta)$ is nondecreasing and continuous with respect to θ. Given any specific probability level $\beta \in (0, 1)$, the β-VaR and β-CVaR values for the loss associated with $\mathbf{x}$ will be denoted by $\theta_\beta(\mathbf{x})$ and $\phi_\beta(\mathbf{x})$.

$$\theta_\beta(\mathbf{x}) = \min\{\theta \in IR : \psi(\mathbf{x}, \theta) \geq \beta\} \tag{38}$$

$$\phi_\beta(\mathbf{x}) = (1-\beta)^{-1} \int_{g(\mathbf{x},\mathbf{p})\leq\theta} g(\mathbf{x}, \mathbf{p}) p(\mathbf{p}) d\mathbf{p} \tag{39}$$

Rockafellar and Uryasev proposed the following convex optimization problem to compute the optimal CVaR portfolio [125].

$$F_\beta(\mathbf{x}, \theta) = \theta + (1-\beta)^{-1} \int_{\mathbf{p}\in R^m} [g(\mathbf{x}, \mathbf{p}) - \theta]^+ p(\mathbf{p}) d\mathbf{p} \tag{40}$$

Rockafellar and Uryasev has proved that the β-CVaR of the loss associated with any $\mathbf{x} \in X$ can be determined from

$$\phi_\beta(\mathbf{x}) = \min_{\theta\in\mathbf{R}} F_\beta(\mathbf{x}, \theta) \tag{41}$$

In this analysis, we apply this method on our model by first approximating $F_\beta(\mathbf{x}, \theta)$ by Monte Carlo simulation [125]

$$\tilde{F}_\beta(\mathbf{x}, \theta) = \theta + \frac{1}{M(1-\beta)} \sum_{j=1}^{M} [g(\mathbf{x}, \mathbf{p}_j) - \theta]^+ \tag{42}$$

where $\mathbf{p_j}$, $j = 1, ..., M$ are vectors sampled from the probability density of the parameters $p(\mathbf{p})$; and then obtain the optimal decision by solving

$$\min_{\mathbf{x}\in X} \phi_\beta(\mathbf{x}) = \min_{(\mathbf{x},\theta)\in X\times\mathbf{R}} F_\beta(\mathbf{x}, \alpha) \tag{43}$$

3.4 Case Study: Ethanol Production by Blue-green Algae

As a case study, we analyze the process design of ethanol production from the blue-green algae process described in [98]. In this technology, ethanol is produced via an intracellular photosynthetic process in cyanobacteria (blue-green algae), excreted through the cell walls, collected from closed photobioreactors as a dilute ethanol-in-water solution, and purified

to fuel grade ethanol. The flow rate out of each bioreactor is significantly smaller than the processing speed of a process unit. Therefore, a large number of bioreactors are needed to achieve a reasonable production. Ethanol can be separated from the ethanol-water mixture produced by bioreactors though distillation and dehydration. Two process designs are proposed 1) vapor compression steam stripping (VCSS) followed by membrane separation, or 2) VCSS followed by distillation and molecular sieve. These two process designs correspond to the two design scenarios developed in section 3.3: VCSS corresponds to local process A, central distillation column corresponds to central process C, and membrane separation corresponds to local process B. A central distillation column can be built in relatively big capacity and has economies of scale, but there is a limit on the size of a VCSS column. Therefore, to take advantage of economies of scale, multiple VCSS columns are built to feed each central process. On the other hand, the cost of membrane separation is expected to be linear with respect to production rate.

Below, we explain the baseline value of the parameters used in this analysis. In order to demonstrate the method used in this paper, we choose a range of approximately 30 % variation around the baseline value for the global sensitivity analysis. For any practical application, the ranges should be chosen based on all the information known to the planner.

Ethanol price: The price of ethanol in the U.S. fluctuated around $2 per gallon in 2009 and rose to about $3.5 per gallon in 2010 [149] [120] [152]. We use a value of $3 per gallon as the average price of ethanol.

Flow rate of each bioreactor: With an expected production rate of 6000 gallons of ethanol per acre per year, the expected flow rate of the ethanol-water solution from each bioreactor is about 300 gallons per month [98].

Bioreactor building speed: The estimated bioreactor building speed is about 500000 bioreactors per year, or about 0.042 million per month [103]. This is equivalent to building an ethanol production capacity of 18 million gallon ethanol per year in a year.

Initial concentration of ethanol: The initial concentration of ethanol is taken to be 1%

[98].

Bioreactor costs: The cost of photobioreactors vary greatly, depending on their size and material [30]. The photobioreactors in the case study are modeled as having a length of 50 feet and a cylindrical diameter of 4 feet, composed of polyethylene of 0.2 mm thickness [98]. The capital cost of each bioreactor is estimated to be around 120 dollars [103].

Bioreactor OPEX: The operating cost of each bioreactor is estimated to be $1.5 per year [103]. With an average flow rate of 300 gallon per month from each bioreactor, this is equivalent to an operating cost of 0.04 dollars per gallon of ethanol.

Benchmark VCSS + Membrane CAPEX: Vane and Alvarez [155] studied a membrane-assisted vapor stripping process with the combination of stream stripper and membrane process to purify 1% ethanol feed to 99.5 % fuel-grade ethanol. They have estimated total capital costs for a 1 million gallon per year ethanol production facility with a 1% ethanol input stream to be $1.3 million, in 2006 dollars, for a stripper operating at or near 65 °C (ca 200 torr) and with a vapor permeation membrane feed pressure of 760 torr. The processing cost of such system for 1% ethanol input stream is estimated to be as low as 0.37 $ per gallon of ethanol produced. Here, we consider the VCSS and membrane separation process as an integrated system and denote its capital cost by C^{AB}. Using this as the baseline case of our analysis, with an average annual inflation rate of 2.35 % from 2006 to 2010, we assume a baseline capital cost of 1.43 million and operating cost of 0.41 dollar per gallon, in 2010 dollars, for a benchmark one MGY facility.

Benchmark VCSS + distillation + molecular sieve CAPEX: Kwiatkowsky et al. [90] have estimated costs for a 40 million gallon per year (MGY) dry-grind corn-based ethanol facility. For the ethanol concentration system, which purifies 10.8 % ethanol-water solution to fuel grade ethanol, consisting of beer column distillation, rectifier, stripper, and molecular sieves, they estimate a system capital cost of $8 million, in 2004 dollars. This value is based on manufacturer price quotes for the components, multiplied by three for system construction costs, again based on discussions with industry. Scaling this value to

the 1 million gallon per year scale with a 0.8 scaling factor results in an estimated $0.42 million capital cost. For a 40 MGY ethanol production rate, the total steam required for beer column, rectifier column and stripper column is about 25 tons per hour, and the operating power for molecular sieve and process condensate tank is about 12 kJ/s. With a steam price of $17.08 per 1000 kg and a electricity price of 0.014 per MJ, this sums up to a processing cost of about $ 0.1 per gallon ethanol produced [90]. With an average annual inflation rate of 2.68 % from 2004 to 2010, we assume a baseline capital cost of 0.49 million and operating cost of 0.12 $ gallon, in 2010 dollars, for an one MGY facility.

Ratios of minimum to maximum flow: There are many different variations of distillation columns. The ratio of minimum to maximum flow depends on the specifics of the particular column. It is commonly believed that this ratio of a sieve tray distillation column is in the range of 0.25 to 1, and a higher turndown ratio is believed to be possible for bubble-cap and valve trays [82] [116]. For the case study, we use a baseline value of 0.5 for both distillation and VCSS columns.

Scaling exponent of production equipment CAPEX: As mentioned earlier, the scaling exponent of bioelectricity plants usually ranges from 0.6 to 0.9 and that of dry mill ethanol plants is estimated to be 0.836 [56] [78]. We take 0.8 as an average scaling exponent.

Benchmark pipeline CAPEX and OPEX: Several major material flows in our case study are transported through pipelines: The dilute ethanol-in-water solution is transported from each photobireactor to the production equipment; water is recycled and piped back to the photobioreactors; carbon dioxide is collected from a near-by industrial source, piped to the production site and inserted to the headspace of photobioreactors for algae growth and ethanol production; and flue gases need to be collected from each photobioreactor and sent to scrubbers where oxygen will be released to the atmosphere and ethanol will be collected. Piping cost is highly dependent on the particular production system and pipeline design. For the purpose of this paper, we estimate piping cost using literature

data and the production system layout of the case study. We consider the pipeline network as an integrated system, which requires one sunk cost and operational energy costs, both of which follow equation 15. For a benchmark processing rate of about 4.6 million gallons ethanol per month, the estimated capital cost of pipeline installation is 5 million dollars and the operating cost per year is about 0.5 million dollars per year. Scaling this value to one million gallon ethanol per year with scaling factor of 1.1 results in an estimated 0.06 million dollars of capital cost. We multiply this number by three to account for all the piping capital cost, which results in 0.18 million dollars CAPEX and about 0.03 dollar per gallon of ethanol produced for a benchmark one MGY facility.

Initial investment cost: The sunk cost at the beginning of the project considered here includes land costs, building costs, administration cost, laboratory cost, etc. We estimate a 10 million dollar sunk cost for a one MGY capacity facility and 80% of this sunk cost is expected to be recovered as salvage value at the end of the project lifetime.

Base operating cost: The base operating cost includes operating cost of production equipment excluding energy costs. The annual maintenance cost (including labor) is assumed to be 7.5% of the cost for installed equipment [155]. The total baseline operating cost is estimated to be 0.10 dollar per gallon.

Energy cost: Energy cost includes heat and electricity consumption required for the operation of production equipment and pipeline operation. Figure 13 shows the estimated energy cost of the integrated VCSS and membrane system as a function of the initial concentration. Obviously, as the initial ethanol concentration increases, the energy requirement for processing the ethanol-water solution into fuel grade ethanol is less.

Project lifetime and the discount factor: The average economic life of the plant is taken to be 20 years [78], with a 10% effective annual interest rate over the 20 years. This is equivalent to about 0.797% compounded monthly.

Table 1 shows the summary of parameter baseline values and lower bound λ_L and upper bound λ_U used in this paper.

Table 1: Parameter baseline value and ranges

	Parameter Description	Baseline	Units
p	Ethanol price	3	\$ /gal
q	flow rate of each bioreactor	300	gal/month
n	Bioreactors building speed	0.042	Million/ month
α	Initial concentration of ethanol	1%	
b	Bioreactor capital cost	120	\$/bioreactor
δ	OPEX of each bioreactor	0.12	\$/month
C_0^{AB}	Benchmark VCSS membrane CAPEX	1.43	M\$
F^A	VCSS fixed cost	0.06	M\$
C_0^A	Benchmark VCSS CAPEX	0.60	M\$
C_0^C	Benchmark distillation CAPEX	0.49	M\$
S_0	Benchmark storage CAPEX	0.1	M \$
γ	ratio of min to max flow	0.5	dimensionless
s	Scaling exponent	0.8	dimensionless
C_0^P	Benchmark pipeline CAPEX	0.2	M\$
C_0^{P1}	Benchmark pipeline CAPEX	0.1	M\$
z	Scaling exponent of piping cost	1.5	dimensionless
e_0	Energy cost parameter	0.0096	dimensionless
u	Energy cost exponent	0.75	dimensionless
C^I	Sunk cost per unit capacity	24	\$ / (gal/month)
o	Base operating cost	0.10	\$ /gal
T	Project lifetime	240	months
d	Discount factor	1.00797	dimensionless

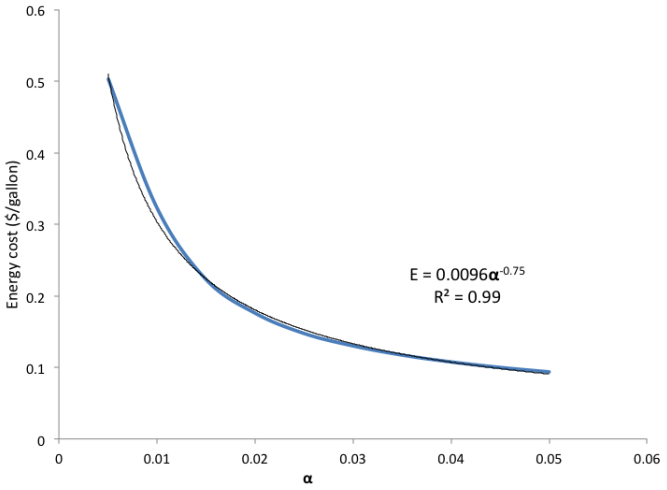


Figure 13: Energy cost as a function of α

3.5 Results and Discussion

3.5.1 Optimal capacity and project value for the integrated VCSS and membrane design

For our reference case, the baseline values presented in table 1 are used for parameter vectors $\mathbf{p_1}$ and $\mathbf{p_2}$ for the integrated and distribute-centralized designs, respectively. Model $y_1 = f_1(x, \mathbf{p_1})$ is solved for the integrated VCSS and membrane separation case. When all parameters in $\mathbf{p_1}$ are realized, y_1 is a strictly concave function of x. Matlab was used to find the optimal value of x, x^*, at which the maximum value of y_1, y_1^*, is realized. Our result shows that, for the reference case, the optimal 1% inlet flow rate to the VCSS + membrane system is $x^* = 2662$ million gallons per month, which is equivalent to an ethanol production capacity of 320 MGY . This 320 MGY facility yields a maximum project value of 1925 million dollars and has a production area of about fifty thousand acres. It is important to notice that this result shows the most economical capacity of a VCSS and membrane combined bioreactor production system. This 320 MGY system can obviously be repeated to achieve a targeted production rate.

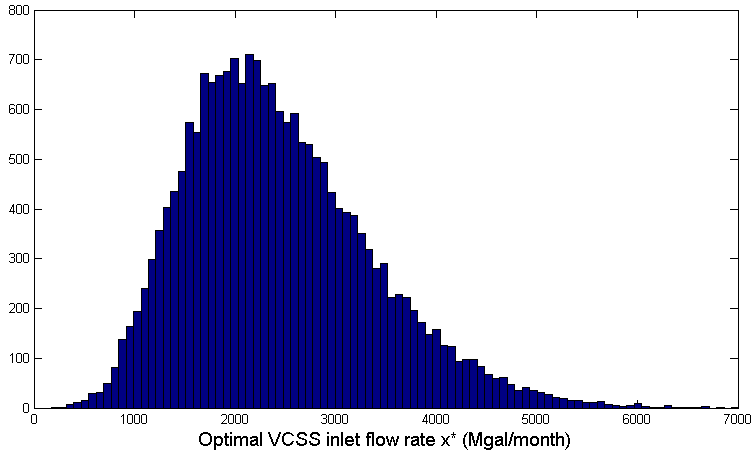


Figure 14: Histogram of the optimal VCSS inlet flow rate x^* based on 100,000 Monte Carlo estimates for each parameter in model $y_1 = f_1(x, \mathbf{p_1})$

3.5.2 Global sensitivity results for the integrated VCSS and membrane design

In order to evaluate the relative importance of the parameters on the decision of the production capacity as well as the project net present value, we calculate the Sobol's indices of all the parameters. All parameters are assumed to be uniformly distributed random variables each within an interval plus and minus 30% of its baseline value. Matlab and R statistical software were used for the global sensitivity analysis. 100,000 Monte Carlo estimates are generated for each parameter. As a result of the variation of parameter values, the optimal production capacity varies around its reference case value, i.e. 2662 Mgal/month. Figure 14 shows the histogram of the optimal production capacity, which is represented by the VCSS inlet flow rate.

Table 2 shows the first order indices and total effect of all parameters on both the decision of production capacity, x, and project net present value, y, for the integrated VCSS and membrane separation option. The first order indices yield the exact fraction of the output variance accounted for by any input parameter or combination of parameters. From the first order Sobol indices for x^*, the results identify 11 important parameters (total first order effect of 0.97) for the optimal capacity decision that cover price of ethanol, bioreactor

outlet flow rate, bioreactor building speed, initial concentration of ethanol from bioreactor outlet flow, turn-down ratio of the production unit, scaling exponent of production equipment, scaling exponent of piping, scaling exponent of energy cost, initial sunk cost, project lifetime and the effective discount factor. Except the turn-down ratio of the production equipment, all these parameters are also the major contributors to the sensitivity of the overall net present value of the project. With the addition of the operating cost factor, these parameters count for a total first order effect of 0.86.

Table 2: Global sensitivity analysis results for model $y_1 = f_1(x, \mathbf{p_1})$

p_j	$S_j(x^*)$	$S_j^{tot}(x^*)$	$S_j(y_1)$	$S_j^{tot}(y_1)$
p	0.09	0.10	0.23	0.28
q	0.26	0.27	0.10	0.13
n	0.18	0.22	0.06	0.09
α	0.02	0.04	0.15	0.20
b	0.00	0.00	0.00	0.01
δ	0.00	0.00	0.00	0.00
C_0^{AB}	0.00	0.00	0.00	0.00
S_0	0.00	0.00	0.00	0.00
γ	0.01	0.01	0.00	0.00
s	0.01	0.01	0.01	0.01
C_0^P	0.00	0.00	0.00	0.00
z	0.08	0.10	0.01	0.02
e_0	0.00	0.01	0.00	0.01
ϵ	0.02	0.02	0.04	0.04
C^I	0.02	0.02	0.00	0.01
o	0.00	0.00	0.00	0.00
T	0.06	0.06	0.05	0.08
d	0.21	0.22	0.21	0.26
Sum	0.97	1.03	0.89	1.08

That neither the sum of first order indices (0.97 for x^* and 0.89 for y_1) is close to one shows that the higher order terms are important. For example, the bioreactor level flow rate q by itself contributes 25 % of the total variance of x^*, and ranks the first in the first order indices of all parameters. However, when the interaction of q with all other parameters are considered, q contributes to 27 % of the total variance of the output x^*. This also means

that if the bioreactor level flow rate is fixed or known, the variance of the optimal size of local process A, x^*, will be greatly reduced. Figure 15 (c) (d) show that x^* and y_1^* both increase close to linearly with respect to the bioreactor building speed n.

For the design of the production system, some of the parameters here can be considered to be either exogenous or endogenous, depending on the project stage at which this problem studied. For instance, capital cost of the production units could be an exogenous factor before the project starts, but it becomes an endogenous factor once the units are purchased or the quote of their price is given. On the other hand, the price of ethanol will always be a stochastic exogenous factor that can not be controlled by the planner. Interestingly, whether to consider the initial concentration of ethanol α as an exogenous or endogenous factor depends on the planner's perspective. If the technology is mature, and this factor is taken as granted, then it can be viewed as an endogenous factor. However, if this factor is a measure of technology advancement, as is the case for the case study considered here, the value of α may increase should the genetically advanced algae be able to produce higher concentrated ethanol-water solution. In this case, α can be viewed as an exogenous factor to study how sensitive is the decision to the possible variation of α. Here, α is considered as an exogenous factor.

The sensitivity results show that α only contributes to 2% of the variance of x^* by itself and 4% of the overall variance when its synergy with all other parameters are considered. And for the optimal net present value y_1, α has a first order index of 0.15 and a total effect of 0.20, ranking the third in the importance of the overall variance in y_1. However, it is important to notice that the sensitivity result for all parameters depend on the range of the values of given as input. Here, the range of α is chosen to be 0.07% to 0.13 %, i.e. 30% plus and minus its baseline value 1%. If the range is increased, the relative importance of α to the overall variances will also increase. If the planner thinks that α has a high chance of increasing in the future, further analysis on the possible range and probability of occurrence of this factor and their impact on the optimal production capacity is necessary. Keeping all

other parameters fixed at baseline value, Figure 15 (e) shows the optimal capacity increases monotonically as initial concentration α increases, and Figure 15 (f) shows that the optimal net present value increases almost linearly with increasing α.

The price factor for ethanol, p, is obviously an important factor for the production systems design. Its total effect contributes to, respectively, 10 % and 28 % of the total variance in x^* and the corresponding y_1^*. As shown in figure 15 (a) and (b), the optimal capacity is a concave function of p within the range selected, and the net present value increases linearly with respect to p.

For the decision of x^*, the total effect of the initial investment C^I contributes to 2 % of the overall variance. A higher initial investment C^I favors a smaller capacity. If the initial investment is too high, the project is not profitable at all, and the optimal plant size is zero. However, intuitively, the initial investment should not affect the sizing of the production system, since it is assumed to be linear with respect to the size of the production rate. One interpretation of this curve is that C^I influences the decision by alleviates the effect of revenue generation of each unit of final product, and therefore reduces the power of time value of money in negotiation with the economies of scale.

Table 3: Global sensitivity higher order Sobol's indices of the VCSS and membrane combined system using model $y_1 = f_1(x, \mathbf{p_1})$

$\mathbf{u_1}$	$S_{\mathbf{u}}(x^*)$	$S^{tot}_{\mathbf{u}}(x^*)$	$S_{\mathbf{u}}(y_1)$	$S^{tot}_{\mathbf{u}}(y_1)$
(q, n)	0.44	0.48	0.16	0.21
(q, n, α)	0.47	0.53	0.32	0.42
(p, q, n, α)	0.56	0.60	0.58	0.65
(p, q, n, α, d)	0.79	0.80	0.84	0.86
(p, q, n, α, z, d)	0.88	0.88	0.86	0.88
$(p, q, n, \alpha, z, T, d)$	0.93	0.93	0.92	0.93
$(p, q, n, \alpha, \gamma, s, z, e_0, \epsilon, C^I, T, d)$	0.98	0.99	0.99	0.99
$(p, q, n, \alpha, b, \delta, \gamma, s, z, e_0, \epsilon, C^I, o, T, d)$	1.00	1.00	0.99	1.00
$(p, q, n, \alpha, b, \delta, C_0^{AB}, S_0, \gamma, s, C_0^P, z, e_0, \epsilon, C^I, o, T, d)$	1.00	1.00	1.00	1.00

Table 3 shows the Sobol's indices of some higher order terms in mode $y_1 = f_1(x, \mathbf{p_1})$. Parameter set $\tilde{\mathbf{u_1}} = \{p, q, n, \alpha, \gamma, s, z, e_0, \epsilon, C^I, o, T, d\}$ can represent the overall variance of both x^* and y_1^*, and set $\tilde{\mathbf{v_1}} = \{b, \delta, C_0^A, S_0, C_0^P\}$ is non-influent. Therefore, we have reduced the model uncertainty set into size of 11. This indicates that a fixed baseline value for $\mathbf{v_1}$ is sufficient for any subsequent analysis.

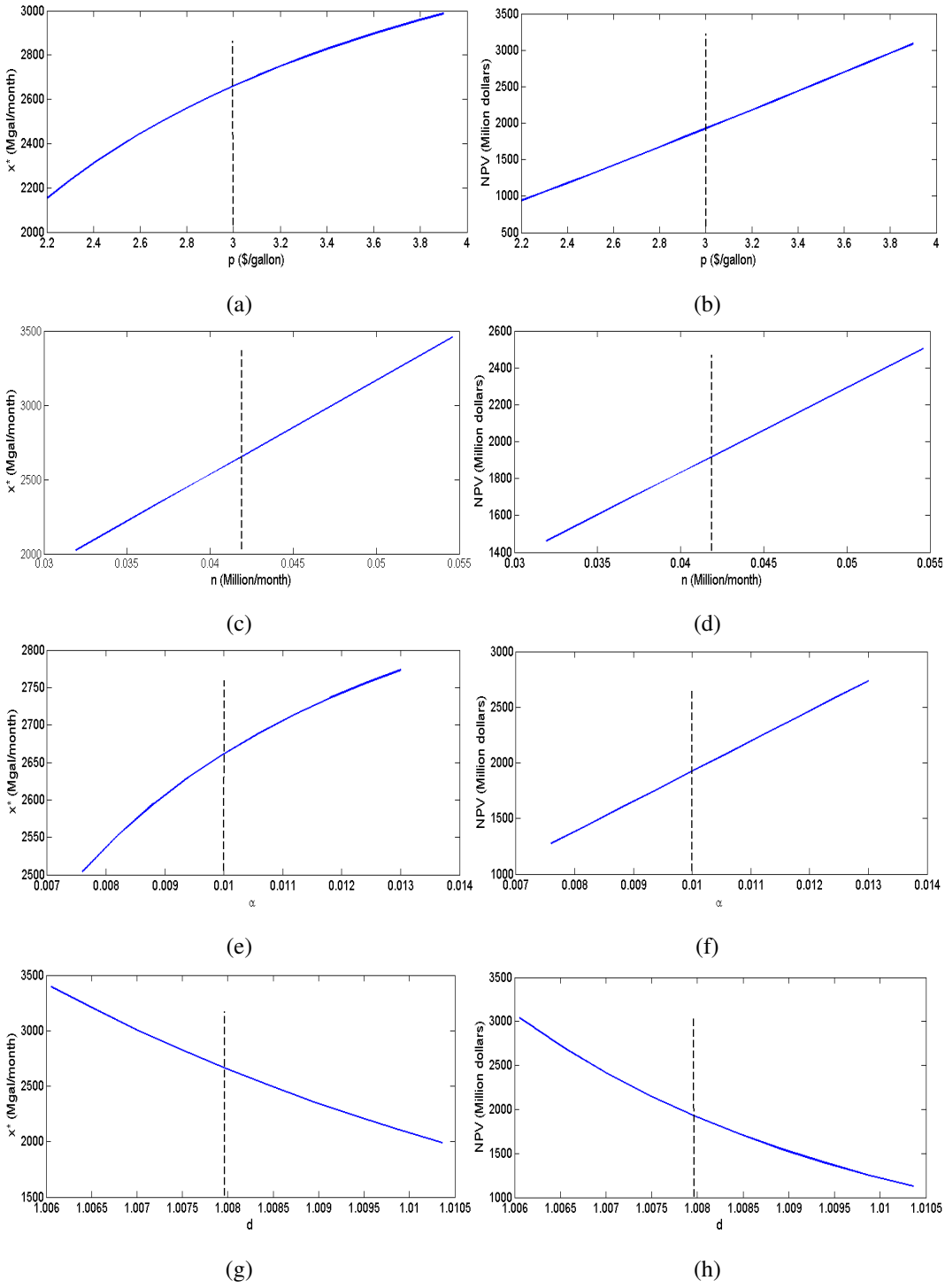


Figure 15: Optimal VCSS inlet flow rate and net present value changes with ethanol price (a) (b), bioreactor building speed (c)(d), initial concentration of ethanol (e)(f) and discount factor (g)(h), while other parameters are fixed. The dashed vertical line shows the baseline value from Table 1

3.5.3 Minimization of CVaR for the integrated VCSS and membrane design

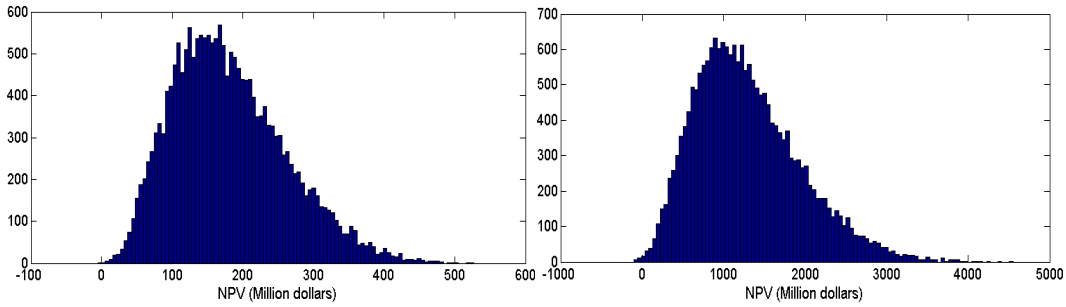
For the reference case, we maximize the net present value with respect to x with the expected value of all the parameters. This is equivalent to maximizing the average outcome of the net present value. A more conservative planner might want to focus on the least profitable scenarios instead. Such a risk averse planner could design the production capacity by minimizing the possible loss. In this section, we consider the case of the VCSS and membrane combined design, and present analytical results of the CVaR minimization problems with loss functions $g_1(\mathbf{x}, \mathbf{p}) = -f_1(\mathbf{x}, \mathbf{p_1})$. Due to the complexity of model $y_1 = f_1(x, \mathbf{p_1})$, the problem can not be solved in closed form. Therefore, given a probability β, Monte Carlo simulation is used to sample $\mathbf{p_1}$ and find:

$$
\begin{aligned}
x^* &= \arg\min_{\mathbf{x}\in X} \phi_\beta(\mathbf{x}) \\
&= \arg\min_{(\mathbf{x},\theta)\in X\times\mathbf{R}} F_\beta(\mathbf{x},\theta) \\
&= \arg\min_{(\mathbf{x},\theta)\in X\times\mathbf{R}} \{\theta + \frac{1}{M(1-\beta)} \sum_{j=1}^{M} [g(\mathbf{x}, \mathbf{p_1}_j) - \theta]^+\}
\end{aligned}
$$

Table 4 shows the optimal x^* selected by minimizing the conditional value at risk of the project with different β. By definition with respect to a specified probability level β, the β-VaR of a portfolio is the lowest amount θ such that, with probability β, the loss will not exceed θ, whereas the β-CVaR is the is the conditional expectation of losses above that amount θ. In this case, for example, a negative 0.95-CVaR indicates that even in the worst 5% cases, the expected net present value of the project is 268 million dollars. The VCSS inlet flow rate that minimizes the worst 5% cases is 1010 million gallons per month, which corresponds to an annual ethanol production capacity of 121 million gallons, about half the optimal production scale of 320 MGY derived in the risk neutral analysis.

Table 4: Results of CVaR optimization for integrated VCSS and membrane design

β	β-VaR (M$)	β-CVaR (M$)	x^* (Mgal/month)	Optimal Capacity (MGY)
90 %	-561	-375	1180	142
94 %	-432	-292	1050	126
95 %	-392	-268	1010	121
98 %	-256	-174	820	98
99 %	-187	-125	710	85



(a) x=100 (b) x=1000

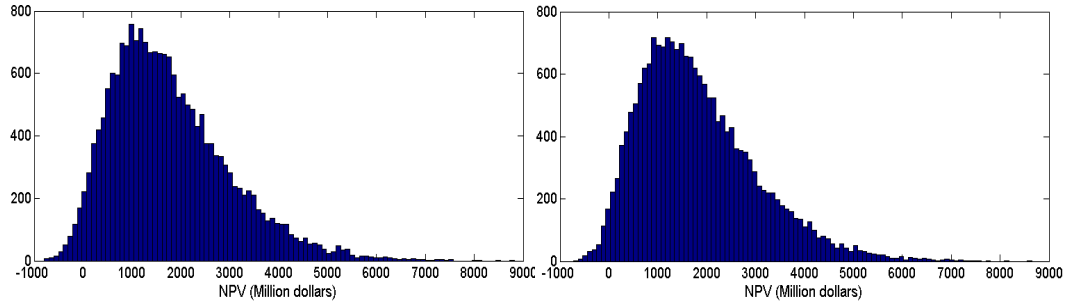
(c) x=5000 (d) x=7000

Figure 16: Histogram of the net present value (million dollars) at different production capacity x (Mgal/month)

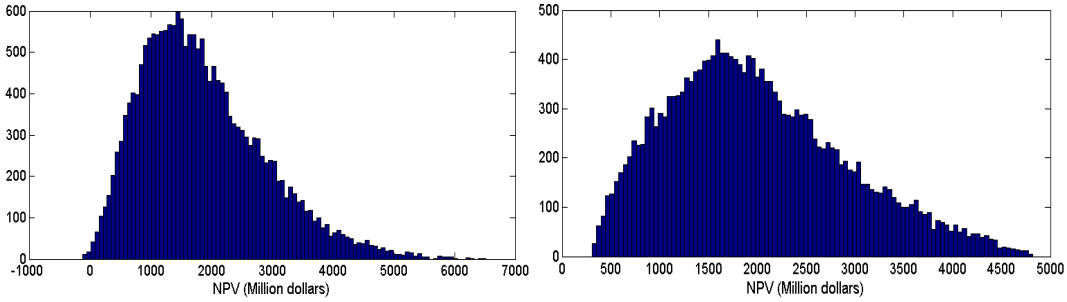
Figure 16 shows the histograms of the net present value at several selected values for x. Even though the shape of the distribution does not change significantly as the production capacity x increases, the range of possible outcomes increases significantly. The coefficient

of variance for the case when the inlet flow rate is 100, 1000, 5000, and 7000 is, respectively, 0.44, 0.50, 1.43, and 7.83. This indicates that a risk averse decision maker will prefer smaller designs to minimize risk. Similarly, if instead of focusing on the leftmost tail of the net present value, we focus on the rightmost tail, i.e. maximize the highest possible profits, then the answer is most likely a higher production capacity. The key lesson here is that the objective function the decision makers choose for designing the size of the system makes a big difference to the optimal x^* they would select. In other words, a risk neutral decision maker who tries to maximize the average profit gained would select a size that could be more than twice the size that risk averse planner would choose.

As mentioned earlier, some parameters considered here can be regarded as endogenous, especially the technological parameters that might be under the control of the design engineer. Once a parameter value in $\tilde{\mathbf{u}}_\mathbf{1} = \{p, q, n, \alpha, \gamma, s, z, C^I, o, T, d\}$ is fixed, the variance of the objective function value will be reduced. Therefore, the decision maker can analyze the parameter values that create the leftmost tail of the distribution and check whether there are any actions that could be taken to reduce the probability that these parameter values will be realized. For example, figure 17 shows how the distribution of the net present value changes as more parameters are realized. As more parameters are realized at baseline value, the tail of the distribution decreases. Figure 17 (d) shows the distribution when only the price of ethanol and the discount factor are kept random. As more parameter values are realized, the difference of the optimal x^* chosen by risk neutral and risk averse decision makers is smaller. This is also intuitively true, since a risk adverse decision maker tend to be more cautious when there are more risk factors present.



(a) $(p, q, n, \alpha, \gamma, s, z, C^I, o, T, d)$ (b) $(p, q, n, \alpha, z, T, d)$



(c) (p, q, n, α, d) (d) (p, d)

Figure 17: Histogram of the net present value at x=2500. The title of each figure indicates the key parameters that are kept random.

3.5.4 Results for distributed VCSS and centralized distillation design

Model $y_2 = f_2(x, k, \mathbf{p_2})$ is tested on the VCSS and central distillation combined design. In this case, several distributed VCSS processes plus one centralized distillation process are allowed. This is motivated by the fact that a distributed VCSS process can reduce the transportation cost by reducing the flow to be transported to the central distillation to one tenth of its original quantity. When the capacity of VCSS is kept unconstrained, solving the problems, $(x^*, k^*) = \arg\max y_2$ for the reference case yields $k^* = 1$ and $x^* = 3950$, which means that the optimal design is to have one VCSS serving one central distillation unit with an ethanol production capacity of 474 MGY. This centralized design result is expected, since it is shown in the analysis of the VCSS and membrane combined system that the dominant factor in capacity sizing is the speed at which the product flow is being built. The equipment cost and piping cost are small compared to the revenue generation

process. Therefore, the effect of the tradeoff between economies of scale in production cost and diseconomies of transportation cost is hidden behind the tradeoff between the drive of generating more revenue and the loss in postponing revenue generation. The postponement in revenue generation is directly influenced by the bioreactor level flow rate, bioreactor building speed and the initial concentration of ethanol in the bioreactor level outlet flow.

Keeping all other parameters fixed at baseline value, solving the net present value maximization model shows that either of the following cases can make distributed VCSS processing preferable to centralized VCSS and distillation process: 1) The scaling exponent of pipeline CAPEX is greater than 1.6; 2) Ethanol price is less than 1.6 dollars per gallon; 3) The effective annual interest rate is higher than 16 %. 4) The min to max flow ratio is higher than 0.8. All these cases enlarge the relative importance of processing cost and transportation cost tradeoff with respect to revenue generation. It is obvious that when the scaling of piping cost becomes steeper, the need to reduce transportation cost is higher. When ethanol price is lower, the project needs to focus on more cost reduction measures, and therefore reducing transportation cost through distributed processing becomes attractive. These results are only an illustration of the effect of parameter values on the centralized and discentralized decision making. Directly interpretating these numerical values can be misleading, since they are all relative to the baseline parameter values.

To further explore the potential of a distributed-centralized process, we calculate the minimum ethanol selling price (MESP), which is defined as the lowest ethanol price at which the net present value of the project is not negative. Table 5 shows the result of the MESP and the corresponding number of local processes supporting a central process under several scenarios. All scenarios are based on the reference case and change in one selected parameter from its baseline value. The first column of table 5 shows selected parameter and the new value that is assigned to it.

The results in table 5 shows that in all the selected scenarios the decentralized design ($k > 1$) is preferred. This indicates that the distributed-centralized design is more robust

Table 5: Minimum ethanol selling price and the corresponding number of VCSS in the system under different scenarios

Scenarios	MESP ($/gal)	k
Baseline	0.93	12
$q = 500$	0.76	10
$n = 0.020$	0.95	10
$\alpha = 0.005$	1.65	5
$\alpha = 0.02$	0.55	16
$\gamma = 0.8$	0.93	19
$s = 0.6$	0.91	8

against low ethanol price compare to a centralized design. As expected, when q, n, α increase, the MESP is lower, since more valuable material stream is generated faster which all have positive impacts on the net present value. Higher ratio of min to max flow favors a more decentralized design, which achieves the minimum flow faster and therefore guarantees an earlier generation of revenue. When the scaling exponent of production equipment s is decreased, the impact of economies of scale is higher and therefore a more centralized design is more favorable.

3.6 Conclusion and Future Research

Biofuels produced from biomass feedstocks, as an alternative to gasoline and diesel fuels derived from petroleum, can provide many benefits, including a reduction in lifecycle greenhouse gas emissions. There are, however, a number of challenges for biofuels to be widely used as a transportation fuel. One of the most important is the production cost, which determines how competitive biofuel can be compared to fossil fuels. Detailed cost analysis and production systems design are important steps towards maximum cost reduction. The models and methods proposed in this paper provide a framework in analyzing biofuel based on highly distributed bioreactor systems, such as algae-based ethanol production.

We use discounted cash flow models for maximizing the net present value of the project

to model the capacity sizing decision. Our model focuses on the postponement in revenue generation that may be caused by the inability of building a large amount of bioreactors required to reach the minimum flow rate of the production equipment. The economies of scale in production cost and possible diseconomies of scale in piping cost are included in the models. We use, as a case study, an algae-based ethanol production system, which requires about 30,000 photobioreactors to reach a one MGY ethanol production capacity. For a separation process involving vapor compression steam stripping and membrane separation, a 320 MGY production capacity yields a maximum project value of about 2 billion dollars, for the reference case. This result, however, is highly influenced by the variations in the parameter values. The global sensitivity analysis identifies, from a total of 18 parameters, 13 parameters that count for all the variations in the result. Bioreactor outlet flow rate, bioreactor building speed and the discount factor are the most influential parameters in capacity sizing. Ethanol price, the discount factor and the initial concentration of ethanol from bioreactor outlet flow are the most influential factors in the project net present value. Using a CVaR optimization approach, we have shown that a risk averse planner who wants to minimize the worst possible scenarios will choose a production capacity that is less than half of the risk neutral planner. Another separation process design involving vapor compression steam stripping and central distillation is studied. Even though the reference case suggests that a centralized design maximizes the net present value, we have shown that a distributed-centralized design is more robust against low ethanol price. When minimum ethanol selling price is used as the objective function instead of net present value of the project, a decentralized VCSS and centralized distillation design is preferred. Global sensitivity analysis on the MESP, as an future research, could be performed to test the parameter influences on the minimum ethanol selling price and the corresponding distributed-centralized design. Even though, the evaluation of a single instance on MESP is relatively slow, the simulation can be easily achieved by parallelization of the tasks, because the computation of each Monte Carlo estimate is completely parallel.

Comparatively, Gallagher et al. [56] estimated an optimal capacity of 60 MGY for dry mill ethanol production facilities by applying statistical analysis on data about existing corn ethanol plant sizes from a USDA cost of production survey. The plant capacity of the 21 ethanol plants in the survey ranged from 9 MGY to more than 90 MGY. Comparatively, for the reference case of our case study, a risk neutral decision maker would choose an optimal capacity of 320 MGY, which is larger than grain ethanol plants, but is close to optimally sized cellulosic biomass plants found to be in the range of 240-486 MGY [169] [10] [56] .

Greenhouse gas emissions (GHG) derived from life-cycle analysis have also been considered as a factor in determining the optimal plant size. Gan and Smith [58] derive formulas for determining the optimal biofuel plant size and the corresponding feedstock supply radius based on the minimization of biofuel production costs less GHG benefits. This can be easily included in the models proposed in this paper. For example, a new parameter g can be introduced as the lifecycle greenhouse gas emission per gallon of ethanol produced. The value of this parameter is studied in [98]. Let p_g be the penalty given to the unit greenhouse gas emission, the price factor p of ethanol can simply be changed to $p - p_g g$ to reflect the environmental cost of the project.

Chapter IV

FORMULATION AND HEURISTIC SOLUTION FOR LOCATION AND ROUTING PROBLEMS OF SOME FIXED AND MOBILE PROCESSING OPTIONS FOR BIOFUELS PRODUCTION

4.1 Introduction

This paper addresses two new types of transportation problems applied in chemical processing industry with linear transportation costs and general production cost structure. In each case, facilities need to be opened to receive flows from a set of predefined sources. The number of facilities, their locations as well as the transportation routes are chosen in order to minimize the aggregated facility and transportation cost function. The primary concern is providing necessary and sufficient conditions for optimal locations and trees on an arbitrary network with respect to this aggregate cost function. This paper explores several different problems occurring in such a context, relates them to the literature and analyzes some new problems. For all problems studied in this analysis, mathematical formulations are presented to formally define the problem, but the focus here is heuristic algorithms that can solve the problems in polynomial time so that sensitivity analysis on the importance of different parameters can be performed.

The focus of this paper is different facility cost structures, especially the impact of concavity in facility cost on the modeling and complexity of the problems, and its practical implications. Here, *facility* refers mostly to production equipment in the chemical industry. The concavity in production cost typically occurs because of economies of scale in production. When inputs to the production units are collected from a large number of point sources, which are geographically dispersed, the production designer may want to decide which equipment size to purchase and where to best place them. The bigger the production

unit, the more point sources are required to support the input flow. When these sources of material flow are highly spatially distributed, higher transportation cost is typically expected. On the other hand, a bigger production unit could take advantage of economies of scale. Such a problem occurs quite frequently in a variety of settings.

Good examples can be found in the biofuel industry, where many of this class of nonlinear facility location problems occur. For instance, satellite processing is expected to reduce cost by allowing some preprocessing of the biomass to be done before transportation to the conversion plant [16]. This preprocessing equipment often has economies of scale in costs and can be regarded as *facilities* to be located. Biomass can be transported by trucks to these *facilities*, the output of which is then transported to a central location to be converted to biofuel. The biofuel producer, instead of collecting biomass from each contracted farmer, may prefer some aggregated regional centers where biomass from the surrounding area can be processed, stored and supplied to the biofuel producer. The biofuel producer, as the decision maker, needs to locate these regional centers considering his overall feedstock supply region. In this case, the economies of scale happen not only in processing cost but also in the ability of the aggregator to provide a more reliable supply because of the large quantities it can handle [16]. Similarly, in the case of biofuel produced from algae grown in bioreactors, the output liquid stream from these bioreactors may need multiple steps of separation in order to reach a desired fuel concentration [98]. Some of the separation steps could be built in small capacity and located close to the bioreactors, so that the liquid stream can be first processed into higher concentrated and more valuable product stream before transported in a long distance to a central location, where the central processing column can be built in much larger capacity to take advantage of economies of scale. Pipeline networks are more common in this case, where intensive liquid and gas transportation are needed. Another interesting example is mobile processing equipment that can be moved around as a vehicle. This vehicle can process feedstocks at each source node and transports the processed feedstock to the central conversion facility. Drying and

densification of corn and cellulosic materials can be done by mobil distributed processors (MPUs) [16]. A recent technology development in algae biomass releases a mobile lysing electromechanical processor that is used to separate the lipids within the algae out of the cells. The electromechanical MPU is run on a dual axle modular trailer that is moved from pond to pond to process algae [60]. The processed algae is then transported by this mobile unit to the conversion facility. In this case, the MPUs are regarded as *facilities* whose routes and sizes have to be decided. The size of a MPU determines the number of sources on its tour.

Even though problems such as satellite processing, bioreactor based fuel process and mobile processing could be achieving exactly the same goal of producing biofuel, the underlying transportation problems are quite different. The first case has been addressed in various contexts in the literature, which is reviewed in the next section. This paper focuses on the latter two cases, and provides mathematical formulations and algorithms for these problems.

We assume that the cost of opening a facility at a feasible site or deploying a mobile distributed processing unit (MPU) may include a *fixed* cost (e.g. building costs, material costs) and a *variable* cost (e.g. energy consumption, labor cost, maintenance). The *variable* cost can be a linear or concave function of the capacity of the facility. Both the *fixed* cost and the *variable* cost are not site-specific. The general cost structure of a facility at site j used in all problems is defined as:

$$f_j(z_j) = \begin{cases} 0 & \text{if} \quad z_j = 0 \\ a_1 + a_2(z_j)^s & \text{if} \quad z_j > 0 \end{cases} \tag{44}$$

where z_j is the capacity of facility located at site j, $a_1 \geq 0$ is the fixed cost, $a_2 \geq 0$ is a constant, and $s \in (0, 1]$ is the scaling exponent. When $s = 1$, production cost is linear with respect to production capacity, and transportation cost is the dominant factor . When $a_2 = 0$, a fixed cost is incurred for every facility chosen. When $0 < s < 1$, unit production

cost decreases with increasing production capacity. In this case, the tradeoff between the economies of scale and transportation costs determine the network. Cases where $a_1 = 0$ or $a_2 = 0$ are also discussed in the following sections.

4.2 Literature Review

Both *facility location problems* (FLP) and *vehicle routing problems* (VRP) are well-known transportation problems in the operations research literature. Detailed explanation can be found in many available review papers [144] [111] [106] [123]. These problems can be modeled as discrete optimization problems. However, even in linear cases, these problem are usually *NP- hard*. Heuristic algorithms and α- approximate algorithms exist, but the latter are rare for nonlinear objectives. Nonlinear facility location problems (FLP) have few references in the literature, but linear FLP problems have been studied much more extensively.

A class of problem where a general facility cost has been studied more extensively is the facility location problem defined on a bipartite graph [69]. From a practical perspective in the biofuel industry, when the material flow is transported through trucking, a direct tour from a source node to a satellite processing site may be more economical. Assume all sources are equivalent in its supply, the required truck tours in one time period are the same. In this case, transportation cost can simply be determined by the distance between a source node and a satellite node. This is a special case of the FLP problem defined on a bipartite graph. A general facility location problem involves finding a set of facilities to serve demands of a set of spatially distributed customers [106] [52], given the distances, times or costs between customers and facilities. Different constraints are applied for different application domains [84] [124]. A generic definition of the problem is described below:

FLP definition Given a bipartite graph $G = (F \cup R, E)$, consisting a set F of potential facility locations, a set R of required source nodes, and and an edge cost E where $\{f, r\} \in E$ indicates that source node r can serve facility at location f, and weight functions $\omega_F \rightarrow \mathbb{N}_{\geq 1}$

and $\omega_E \rightarrow \mathbb{N}_{\geq 1}$ and $k \in \mathbb{N}$. Find a set $F^{'} \subseteq F$ of facility locations and a subset $E^{'} \subset E$ of edges such that 1) $\forall f \in F(f \in F^{'} \iff \exists e \in E^{'}(f \in e))$; 2) $\forall r \in R(\exists e \in E^{'}(r \in e))$ 3) $\sum_{f \in F^{'}} \omega_F(f) + \sum_{e \in E^{'}} \omega_E(e) \leq k$.

A p-median problem, as the simplest form of the class of FLP problem, selects p facilities to minimize the total distance for serving the customers [69] [43] [123], i.e. $\mid F^{'} \mid = p$. In this paper, however, we are especially interested in the case where the number of facilities to be located is an endogenous decision. This is similar to a class of problem called uncapacitated facility location problem (UFLP) [124] [107], which is a well-known NP-hard combinatorial optimization problem. The general problem of facility location problem with uniform demand can be formulated as follows:

$$\min \sum_{(i,j) \in E} c_{ij} x_{ij} + \sum_j f_j(z_j) \tag{45}$$

$$\sum_j x_{ij} = 1 \quad \forall i \in R \tag{46}$$

$$\sum_{(i,j) \in E} x_{ij} \leq z_j \quad \forall j \in F \tag{47}$$

$$x_{ij} \in (0,1); z_j \geq 0 \tag{48}$$

where c_{ij} is the cost of traversing edge (i, j), x_{ij} is the indicator for whether edge (i, j) is chosen, z_j is the number of sources that serves a facility and $f_j(z_j)$ is defined by equation 44. If $f_j(z_j)$ is convex or if $s = 1$ and $a_1 = 0$, the problem can be easily solved by sorting all the edges in E and selecting the smallest $\mid R \mid$ edges that satisfies the constraint 46. Several studies have provided algorithms for solving the case when $s = 1$ and $a_1 > 0$ [144].

Concave facility cost structure ($s \in (0, 1)$ and $a_1 > 0$) of the above formulation has also been addressed in the literature [177] [36] [69] and it is shown that the concavity of the cost curve due to economies of scale leads to multiple-optima [51]. Hajiaghayi et al [69],

motivated by an application in placing servers on the Internet, provided a greedy algorithm with an approximation factor 1.861 for the case of unit customer demand. Feldman et al [51], for example, used a type of heuristic based on local volume approximation. A local customer set for each facility is defined by choosing those customers to whom the facility is closest based on transportation cost; the sum of demand in this local customer set is assigned to the facility and is used to estimate the cost of this facility. The heuristic then examines the cost of supplying a given customer from each of the available facilities with their cost estimated from their total local demand. Once a local volume is established for each candidate *facility*, the "drop" approach eliminate them one by one as long as a cost reduction can be achieved. Many refinements of the add and drop algorithm were proposed later. Broek, P. and Schutz [25] solved the problem to optimal with Lagrangian relaxation but by approximating the concave function with piece-wise linear, non-convex, non-concave functions. An iterative procedure is proposed in [177] based on tangent line approximation [83].

Many considerations in the chemical processing and biofuel industry can be mapped onto the FLP literature. Large economies of scale in biomass processing usually require a large area of biomass feedstock supply, and therefore the tradeoff between transportation cost and production economies of scale plays an important role in the economic viability of the plant [22]. Biomass feedstocks collected from the sources may need multiple steps of processing to become the final product. Some pre-processing technologies have been developed that process biomass into denser energy carrier or intermediate products, such as bio-oil or bio-slurry, before upgrading to liquid transportation fuels [154] [101]. Distributed preprocessing on site or in a fixed preprocessing facility can provide significant cost benefits by improving handling, transporting, and merchandising potential [168]. These preprocessing sites can be located all in one place or become part of the location decision model. In the latter case, the problem is similar to the multi-layer supply chain problem in the FLP literature [127] [126], and the preprocessing sites, as hubs, provide

potential to reduce transportation costs by processing biomass to a more valuable dense feedstock.

Several recent contributions in the literature have addressed the production economies of scale in the biofuel industry and the potential of using a distributed-centralized biofuel production layout with optimization (MILP) models. Bowling et al [22] considers the decision whether to adopt distributed or centralized processing of biomass facilities with a mixed integer linear programming (MILP) formulation, in which they maximize the total net profit considering the product sales, cost of raw material, transportation costs and equipment capital and operating costs. They have shown that distributed configurations usually represent better solutions than centralized solutions. Corsano et al [39] proposed a mixed integer nonlinear programming model for the design and behavior analysis of the sugar/ethanol supply chain. Akgul et al [3] presented MILP model to optimize the locations and scales of the bioethanol production plants, biomass and bioethanol flows between regions, and the number of transport units required for the transfer of these products between regions as well as for local delivery.

Optimization of the biofuel supply chain under multiple performance measures has also been addressed. You et al. [176] present a multi-period, multi-objective MILP model, which minimizes simultaneously the total annual cost and lifecycle GHG emissions of the cellulosic biofuel supply chain. Their model takes into consideration many characteristics of the biofuel supply chain, such as biomass deterioration with time, geographical diversity, moisture content, etc. Zamboni et al [178] and Mele et al [104] also presented models that consider both economic and environmental aspects of the bioethanol and sugar production.

Here, we focus on the tradeoff between transportation cost and production economies of scale. And different from the literature mentioned above, which are all based on facility location on bipartite graph, we consider, in section 4.3, the case of pipeline construction which allows connect between not only source and sink nodes but also between source nodes; and, in section 4.4, we consider the case of mobile distributed processing.

4.3 Facility location and sizing with underlying pipeline network

In this section, we consider integrated production at facility locations that collect flow from an underlying pipeline network. We are interested in the case where transportation cost is so high that opening another *facility* might be more cost beneficial. Given a set of facilities at specific locations and the regions they are serving, the transportation network is the minimum total distance to connect all the point sources in its region to each facility. Each point source should be serving at least one facility. In the example shown in figure 18, all the *required nodes* are connected but not all the possible facility locations. If there is only one candidate *facility* location, chemical flow has to be transported from point sources to this *facility* on a continuous basis, and the problem is reduced to a *minimum spanning tree* problem.

The problems in this section are based on the following assumptions:

- Each source node has unit supply of material flow;
- It is possible to build a facility that is big enough to process flow from all source nodes, i.e. there is no capacity constraint on the total flow processed by a facility;
- A facility is served by a source as long as there is a path from the facility to the source node.

We have assumed that the production units are uncapacitated here in order to find the theoretically optimal capacities for the selected locations. This is obviously not true in practice, since an infinitely large production unit can not be built. However, it is a valid assumption in some cases. For example, a very large distillation column can usually be built that is enough to process all input flow collected from the source nodes.

Definition 1 ***Facility location and sizing with pipeline network (FLSPN)*** *Given an undirected graph $G = (V, E)$, and an edge cost $\omega_E \rightarrow R^+$. V can be partitioned into two disjoint*

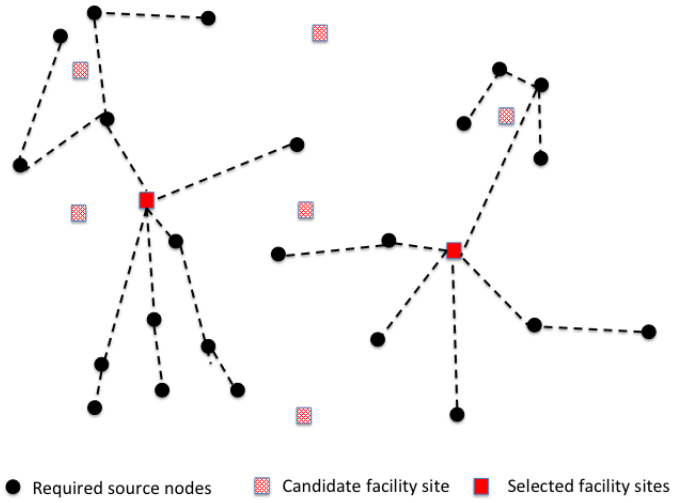


Figure 18: A simple example with two facilities

sets $V = R \cup F$, where $R \subseteq V$ is the set of required nodes and $F \subseteq V$ is the set of possible facility sites. Let the cost of opening a facility of capacity z_j at location $j \in F$ be $f_j(z_j)$ defined by 44. The problem is to choose a subset $F' \subset F$ of facility locations and a subset $E' \subset E$ of edges such that 1) $\forall f \in F(f \in F' \iff \exists e \in E'(f \in e))$; 2) $\forall r \in R(\exists e \in E'(r \in e))$ 3) The total cost $\sum_{j \in F'} f_j(z_j) + \sum_{e \in E'} \omega_E(e)$ is minimized.

Figure 18 shows a simple example where the 23 *required nodes* (black) are covered by two *facilities*, having a capacity of 13 and 10, respectively. There are, in total, eight possible locations where a facility can be opened. To model the facility location problem, we define two binary variables below:

We formulate the generic problem as a network flow problem, which allows us to capture the total flow into a facility. In order to do so, we construct a directed graph by replacing each edge $(i, j) \in E$ with two directed edges (i, j) and (j, i) in the opposite way and construct the new graph $\tilde{G} = (V, \tilde{E})$. Let,

$$x_{ij} = \begin{cases} 1 & \text{if edge } (i, j) \text{ is chosen} \\ 0 & \text{otherwise} \end{cases}$$

$y_{ij} =$ Flow on edge (i, j)

Let the problem be formulated as

$$min \sum_{(i,j)\in\tilde{E}} c_{ij}x_{ij} + \sum_{j\in F} f_j(z_j) \tag{49}$$

subject to,

$$\sum_{k\in V} y_{jk} - \sum_{i\in V} y_{ij} = \begin{cases} 1 & \forall i \in R \\ -z_k & \forall j \in F \end{cases} \tag{50}$$

$$0 \le y_{ij} \le N x_{ij} \quad \forall (i,j) \in \tilde{E} \tag{51}$$

$$x_{ij} \in \{0,1\} \quad \forall (i,j) \in \tilde{E} \tag{52}$$

$$z_j \ge 0 \quad \forall j \in F \tag{53}$$

Here, constraint 51 ensures that flow is only sent on an edge if the edge itself is selected. Constraint 50 guarantees that each source node is connected and its flow is sent and each chosen facility receives a positive flow.

This formulation can easily handle several side constraints and variations. If the source nodes have different supply of flow, then constraint 50 can be easily modified to the amount of supply at each node instead of one. If the cost of an edge is not constant but rather a function of the amount of flow on it, then c_{ij} can be modified to $c_{ij}(y_{ij})$. A capacity limit of facilities can be added by adding an upper bound to z_j.

Below, we discuss three variations of the problems defined above that result in different structure of the facility cost.

4.3.1 $FLSPN1$ problem with zero fixed production cost and linear variable production cost

When the facility cost is linear, i.e. $s = 1$ and $a_1 = 0$, the number of facilities chosen does not affect the objective function value. This is because the total facility cost is fixed

at $C_F = a_2 N$, where N is the total number of point sources. In this case, the problem is reduced to minimizing the total transportation cost under the constraint that each point source is connected with at least one *facility*. In an extreme case, where there is a candidate facility located at every point source, the total transportation cost is zero, and the optimal solution is to open a production unit at each point source. We define the problem below

Definition 2 ***FLSPN1: one-step production with*** $s = 1$ ***and*** $a_1 = 0$ *Given an undirected complete graph* $G = (V, E)$*, where* V *can be partitioned into two disjoint sets* $V = R \cup F$*, where* $R \subseteq V$ *is the set of required nodes and* $F \subseteq V$ *is the set of facilities; and a symmetric nonnegative edge cost function* $c_E \rightarrow \mathbb{R}_{\geq 1} : c_{ij} = c_{ji}$*. The edge cost satisfies triangle inequality. Find a minimum cost subgraph that touches all the required nodes and such that there is a path from each node in* R *to a facility in* F*.*

Notice that even though we formulate the problem as a network flow problem, this variation bears more similarities with the Steiner Forest Problem in that it finds a set of trees that spans all *required nodes* with the option of including some other nodes on the graph. The difference of the FLSPN1 problem with Steiner Forest Problem is two-fold: 1) the subsets of *required nodes* that need to be connected with each other are not pre-defined; 2) in each resulting tree of the forest there has to be one facility.

Lemma 1 *FLSPN1 problem is easy.*

We will prove that the FLSPN1 problem in P by providing the following algorithm that can be solved in polynomial time. Without loss of generality, assume $|R| = n$, $|F| = m$, and $|E| = k$.

Algorithm $FLSPN1$

Initialization: $FLSPN1 = \emptyset$

Step 1: Sort all edges in E in non-decreasing c_{ij} in a list $LIST$, and let $T = \{\}$ and $count = 0$;

Step 2:

while (Not all nodes in R are in $FLSPN1$) OR ($|FLSPN1| \leq |R| + count - 1$) **do**

Remove (i, j) from the top of $LIST$;

if Adding (i, j) to $FLSPN1$ does NOT create a cycle **then**

Add (i, j) to $FLSPN1$

IF $i \in F$, THEN $T = \{T, i\}$ and $count = count + 1$

else

Discard (i, j)

end if

end while

Step 3: If $count = 0$, scan the next (i, j) from the top of $LIST$;

if (i, j) has one node in R and another node in F **then**

Add (i, j) to $FLSPN1$ an remove (i, j) from $LIST$

else

Do nothing

end if

Step 4:

for k=1 to *count* **do**

Find the next largest edge (i, j) in $FLSPN1$

if After deleting (i, j), there is both a path from i to a node in T and a path from j to a node in T, i.e. it results in no tree that contains no node in T **then**

Delete (i, j) from $FLSPN1$

else

Do nothing.

end if

end for

Claim 1 *Algorithm FLSPN1 is a polynomial algorithm.*

Proof 1 *There are four primary steps in Algorithm FLSPN1:*

- *Step 1: Sort the edges:* $O(k \log k)$*;*
- *Step 2: Store the edges in a forest structure, scan LIST and add* (i, j) *to FLSPN1. Each scan takes* $O(m + n)$ *and k potential arcs takes* $O(k)$*;*
- *Step 3: Scan edges in LIST takes at most* $O(k)$*.*
- *Step 4: This is done by setting* $list_0 = \{i\}$*, and while* $list_0$ *is not empty, pick a node from list and check all edges* (i, x) *going out of it, if* $x = j$ *for any of those edges, stop; otherwise add x to the list. This step takes at most* $O(k)$*.*

Therefore, the overall algorithm takes $O(k \log k + m + n + k) = O(k \log k + m + n)$

Claim 2 *Algorithm FLSPN1 solves the FLSPN1 problem to optimality.*

Proof 2 *First, we prove that Algorithm FLSPN1 produces a subgraph that spans all required nodes and contains at least one* $f \in F$ *in each connected part of FLSPN1*

- *Let P be the subgraph created by Step 2 of Algorithm FLSPN1. Then P cannot have a cycle because if an edge creates a cycle, it would not have been added to P. And P is a spanning tree because at the end of Step 2, there are* $|R| + |T|$ *nodes in P and exactly* $|R| + |T| - 1$ *edges without cycle. Therefore, P is a subgraph that spans all nodes in R*
- *Step 3 makes sure that at least one* $f \in F$ *is included in P;*
- *Step 4 deletes edges while making sure that each resulting tree contains at least one* $f \in F$

Next we prove that the solution by Algorithm $FLSPN1$ can not be improved, by assuming first that there exists another subgraph Y that satisfies all the requirements and has less total cost than $FLSPN1$ and prove that it is not possible.

- *Assume that there exists another subgraph Y that satisfy all the requirements and has less total cost than $FLSPN1$.*
- *It is obvious that Y has at least n edges;*
- *$FLSPN1$ created by Algorithm $FLSPN1$ has exactly n edges that are shortest in the $LIST$. Replacing any other edge in $FLSPN1$ will either result in a cycle or a tree without a facility.*
- *Therefore, the n edges in $FLSPN1$ must have a total cost at least as small as the total edge cost in Y, which proves the assumption untrue.*

Therefore, Algorithm $FLSPN1$ produces an optimal solution to the problem $FLSPN1$.

Therefore, we have just proved that $FLSPN1$ problem is easy and can be solved in polynomial time by Algorithm $FLSPN1$.

4.3.2 $FLSPN2$ problem with positive fixed production cost and linear variable production cost

It is common that a fixed cost exists whenever another production equipment (*facility*) is purchased regardless of its size. When a fixed production cost exists, the facility cost $f_j(z_j)$ can not be ignored as in $FLSPN1$. In this case, bigger optimal facility sizes are expected. We do not want to incur another fixed cost unless it can reduce transportation cost more than the fixed cost itself. However, $s = 1$, the linear variable production cost part can be neglected. This is because the objective function defined by equation 49 can be reduced as follows:

$$\sum_{(i,j)\in E} c_{ij}x_{ij} + \sum_{j\in F} f_j(z_j) = \sum_{(i,j)\in E} c_{ij}x_{ij} + \sum_{j\in F}(a_1 + a_2 z_j) = \sum_{(i,j)\in E} c_{ij}x_{ij} + x_m a_1 + Na_2 \qquad (54)$$

As in the $FLSPN1$ problem, the term Na_2 is fixed regardless of the decision. Therefore, the problem is equivalent to minimizing $\sum_{(i,j)\in E} c_{ij}x_{ij} + x_m a_1$ where x_m is the number of facilities included in the solution.

Definition 3 ***FLSPN2: one-step production with*** $s = 1$ ***and*** $a_1 > 0$ *Given an undirected complete graph $G = (V, E)$, where V can be partitioned into two disjoint sets $V = R \cup F$, where $R \subseteq V$ is the set of required nodes and $F \subseteq V$ is the set of facility; and a symmetric nonnegative edge cost function $c_E \rightarrow \mathbb{R}_{\geq 1} : c_{ij} = c_{ji}$ and a fixed facility cost $a_1 > 0$. The edge cost satisfies triangle inequality. Find a set of trees that spans all the required nodes and each contains at least one facility such that the total edge cost and fixed facility cost are minimized.*

Lemma 2 *FLSPN2 problem is NP-hard.*

We will prove this by reducing a well-known NP-hard problem–the Vertex Cover Problem– to the $FLSPN2$ problem. A vertex cover of a graph is a set of vertices such that each edge of the graph is incident to at least one vertex of the set. Given a graph $G = (V, E)$, the **Minimum Vertex Cover Problem** finds the smallest number k such that G has a vertex cover of size k.

Given an instance of a vertex cover problem, defined by a graph $G = (V, E)$, we now construct a $FLSPN2$ instance specifying the graph $V = (G, E)$, the set R, the set S, and edge cost ω_E.

- For a graph $G = (V, E)$ of a Vertex Cover Problem where $|V| = m$ and $|E| = n$, we construct a new graph $G' = (V', E')$ that has $(m + n)$ vertices.
- Define the set of vertices V': for each vertex v in G construct a vertex $[v]$ in G', and for each edge (u, v) in G construct a vertex $[uv]$ in G'.
- Define the set of required nodes $R \subset V'$: $R = \{[uv] : (u, v) \in E\}$

- Define the set of facilities $F \subset V^{'}$: $F = \{[v] : v \in V\}$
- Construct the set of edges: $E^{'}$ contains edges between each pair of vertices in $V^{'}$, i.e. the graph is complete;
- Define edge costs:
  - For every edge $(u, v) \in E$, the edge cost between $[u]$ and $[uv]$ is one and between $[v]$ and $[uv]$ is also one;
  - For every two vertices $u, v \in V$, the edge cost between $[u]$ and $[v]$ is one;
  - For $u \in V$ and $(v, l) \in E$ where $u \neq v \neq z$, the edge cost between $[u]$ and $[vl]$ is two;
  - For every two edges $e_1, e_2 \in E$, the edge cost between $[e_1]$ and $[e_2]$ is two if e_1 and e_2 share an endpoint and three if they are disjoint.
- Define node costs: for every node v in V, assign cost one to $[v]$; for every edge (u, v) in E, assign cost zero to node $[uv]$.

It's easy to see that the reduction from Vertex Cover Problem to $FLS\,PN2$ problem can be done in polynomial time. Figure 19 shows an example of the graph constructed by this transformation. The construction of the edge cost guarantees *triangle inequality*, so the graph is defined on metric space.

Claim 3 *If there is a vertex cover in G with k vertices, then there is a FLS PN2 subgraph in $G^{'}$ of cost n + k; and if there exists a FLS PN2 subgraph of $G^{'}$ of cost $\leq (n + k)$, then there is a vertex cover for G of size k*

Proof 3
- $VC \Rightarrow FLS\,PN2$

  Suppose there is a vertex cover C for G of size k. Clearly, C covers all the edges in E. For each node $v \in C$, select $[v] \in G^{'}$; and for each edge uv with $[v]$ as an endpoint,

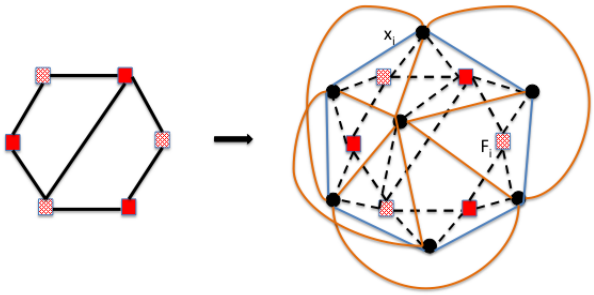


Figure 19: Transformation from an instance of the Vertex Cover Problem to an instance of $FLSPN1$ problem

select edge $([v],[uv]) \in E^{'}$*, the resulting subgraph in* $G^{'}$ *contains k facility nodes and all required nodes and has a cost of* $n+k$*. than k edges.*

- $FLSPN2 \Rightarrow VC$

  Suppose now there exists a $FLSPN2$ *subgraph with cost* $\leq (n+k)$*. For facility nodes* $[v]$ *in the subgraph, include the corresponding v into a set C. Since the subgraph has to cover all required nodes, the nodes in C cover all the edges in G. Since in order to cover all required nodes in* $G^{'}$*, there has to be at least n edges, there are at most k facility nodes in the subgraph. Therefore,* $|C| \leq k$*, which indicates that there must exist a vertex cover of size k.*

This concludes the proof.

Since the vertex cover problem is NP-hard, and we have proved that the vertex cover problem can be reduced to the $FLSPN2$ problem, then $FLSPN2$ problem must NP-hard [59].

Below, we provide a polynomial algorithm for the $FLSPN2$ problem:

Algorithm $FLSPN2$

Step 1: Let $T \rightarrow \{\}$

for $f \in F$ **do**

Find a minimum spanning tree t with {R, f}

$T \rightarrow \{T, t\}$

end for

Let $t_1 = \arg\min(c_t \mid t \in T)$ and let $\mathbf{Q} \rightarrow \{t_1\}$

Step 4: Choose a tree t_i in $\mathbf{Q}$, let k_i the number of *required nodes* in t_i and f_i be the facility in t_i

for $f \in F \setminus \{f \mid f \in \mathbf{Q}\}$ **do**

find a minimum spanning tree t_2 on set $\{R, f_i, f\}$

find the longest edge in this tree $e_{max} \in t_2$

if Deleting e_{max} results in two trees each of which contains one *facility* **then**

Delete e_{max}, and re-optimize the two new trees t_3 and t_4

Record the number of *required nodes* in both trees k_3 and k_4, respectively

else

Find the next longest edge and repeat

end if

if $C(t_i) - C(t_3) - C(t_4) > f(t_3) + f(t_4) - f(t_i)$ **then**

$\mathbf{Q} \leftarrow \{\mathbf{Q}, t_3, t_4\} \setminus \{t_i\}$

else

Do nothing

end if

end for

Step 5: For each tree in $\mathbf{Q}$, repeat Step 4;

Step 6: Output $\mathbf{Q}$

Therefore, $\mathbf{Q}$ is a solution to the $FLSPN2$ problem

4.3.3 $FLSPN3$ problem with positive fixed production cost and concave variable production cost

Definition 4 ***FLSPN3: one-step production with*** $s < 1$ ***and*** $a_1 > 0$ *Given an undirected complete graph* $G = (V, E)$*, where* V *can be partitioned into two disjoint sets* $V = R \cup F$*, where* $R \subseteq V$ *is the set of required nodes and* $F \subseteq V$ *is the set of facility; and a symmetric nonnegative edge cost function* $c_E \rightarrow \mathbb{R}_{\geq 1} : c_{ij} = c_{ji}$ *and a facility cost* $f_j(z_j) = a_1 + a_2(z_j)^s$ *where* z_j *is the capacity of the facility,* $s < 1$*,* $a_1 > 0$ *and* $a_2 \geq 0$*. The edge cost satisfies triangle inequality. Find a set of trees that spans all the required nodes and each contains at least one facility such that the total cost of the trees and total facility cost are minimized.*

Proposition 1 *$FLSPN2$ Algorithm provides a feasible solution and upper bound to the $FLSPN3$ problem.*

The proof is straightforward. Algorithm $FLSPN2$ produces a set of trees that spans all *required nodes* and each tree contains a *facility*. This, by definition, is a solution to the $FLSPN3$ problem. Given a subgraph **Q**, which defines a set of *facilities* of which *facility* j is served by z_j source nodes, $\sum_j (a_1 + a_2 z_j) \geq \sum_j (a_1 + a_2 (z_j)^s) \quad \forall s \in (0, 1)$. Therefore the cost of **Q** is an upper bound on the optimal solution of the $FLSPN3$ problem.

4.4 Mobile distributed processing and variants of the Vehicle Routing Problem

When there is a mobile processing unit (MPU) that can be moved to each bioreactor or pond to process the material, the problem becomes a vehicle routing problem (VRP) with the MPU as the ”vehicle”. A basic vehicle routing problem can be simply defined as follows: Given a set of customer nodes and some vehicles located a single depot node, VRP finds a set of vehicle tours that all start and end at the depot, such that each customer node is visited exactly once and the total cost is minimized. Similarly, given the supply and location of the source nodes, the planner wants to know how many MPUs (vehicles) are needed to process

flows for all the sources and which route each mobile unit takes to serve them such that the total transportation cost and the cost of the MPUs are minimized.

There are many variations that can result in this planning problem. Most of the time, the output of the mobile units needs to be transported to a location where further processing can be done. There can be one or many such final locations to which the MPUs travel. These final locations serve as *depots* in vehicle routing problems. It is uncommon but possible that the mobile units complete all the processing needed, in which case, they do not have to return to any *depot* and the problem is reduced to finding a set of tours on the required source nodes. Generally speaking, the specific problem considered in this paper is a *capacitated vehicle routing problem with non-homogeneous fleet*: There are K vehicles with different capacities located at a single depot node. Given a set of source nodes and the amount to be picked up at each node, find a set of vehicle tours that all start and end at the depot, such that each source node is visited exactly once and the total transportation and vehicle cost is minimized.

As before, we are interested in the capacity design of the MPUs, the cost of which are determined by their capacity. For the facility location problems discussed in previous sections, we have assumed that the production units are uncapacitated in order to find the theoretically optimal capacities for the selected locations. For mobile distributed processing, however, if all MPUs are assumed to have infinite capacity, all cases considered in this paper become a traveling salesman problem (TSP), i.e. finding one optimal tour that visits all source nodes. The reason is that when no other constraint exists, by combining two tours into one, both transportation cost and facility cost can be reduced, which results in one single TSP tour for all source nodes. Since TSP itself is NP-hard, all cases considered here would still be hard. However, this case is not particularly interesting, since it is less common to build a very large MPU, especially considering the fact that a mobile unit most likely has a maximum weight limit to be moved around. Moreover, the time a MPU spends on its tour increases linearly with the total length of that tour, which increases as more

source nodes being served by one MPU. The bigger the MPUs are, the more material flow have to be put in storage at each source node before the MPU visit the node to process the flow, which may incur higher storage cost. If the material flow is biomass, longer storage time may result in biomass degradation and decreasing yield. On the other hand, bigger MPUs also require bigger storage space at the depot where further production is done, which also increase costs. Therefore, very large MPUs are not economical and hence are not preferred. As a result, we consider the case where only a limited number of predefined types of MPUs are available. Each type of MPU has a specific capacity. However, in order to keep focusing on capacity design, we allow a large number of capacity types and a large number of vehicles of each type.

The cost of a MPU (vehicle), same as production unit (facility) in previous sections, contains a fixed cost and a variable cost that depends on its capacity. Since the capacities of the vehicles are predefined, their corresponding costs are predefined as well. However, knowledge of the cost structure of a vehicle could give more information in algorithm design. To our knowledge, the cost structure of vehicles is rarely mentioned in the literature. In most VRP literature, the cost is assumed to be proportional to the distance traveled by vehicles with a few exceptions. Anily and Federgruen [5] studied the capacitated routing problem where the cost of a route is dependent on both the total distance traveled and the number of source nodes on the route.

In the previous sections, we have specifically considered the case where all source nodes have the same flow rate. Here, we relax this assumption to consider source nodes with unsplit non-homogenous demand.

4.4.1 Problem description and formulation

This section presents a formal definition and a mathematical formulation of the mobile distributed processing problem. The MDPP is defined as follows:

Definition 5 ***Mobile distributed processing problem (MDPP)*** *Given an undirected complete graph $G = (V, E)$, where $V = \{v_0, v_1 ..., v_n\}$ and an edge cost $E \rightarrow R^+$. v_0 corresponds to the depot, and $\{v_1, ...v_n\}$ correspond to the source nodes. The edge cost satisfies the triangle inequality and $c_{ij} \neq c_{ji}$. Each source node i (i=1,...,n) is associated with a known nonnegative supply, d_i, to be processed and transported by the vehicle to the depot. A set of K non-identical vehicles are available at the depot. Vehicle k has a capacity of Z_k and a nonnegative cost f_k. The MDPP consists of finding a set of tours (vehicle routes) with minimum cost, defined as the sum of the costs of the edges belonging to the tours, and such that*

- *each tour starts and ends at the depot v_0*
- *each source nodes is visited by exactly one tour; and*
- *the sum of the supply of the source nodes visited by a vehicle k does not exceed the vehicle capacity, Z_k*

Below we formulate the generic problem by modifying a vehicle flow problem formulation by Christofides [34]. By using an index that counts the number of times a particular type of vehicle traverses an edge, we are able to count the number of source nodes and the total supply a vehicle is receiving. This allows us to incorporate the cost of vehicles in the objective function.Given K vehicles, each of capacity Z_k. Let,

- l_e be the transportation distance on edge e
- c_e be the cost of traversing edge e, which is defined as a linear function of l_e ;
- f_k be the cost of vehicle k, which is defined as a concave function of its capacity Z_k following equation 44;
- d_i be the demand at source node i;

And we define the decision variables as follows:

- $x_{ek} =$ the number of times edge e is traversed by vehicle k;
- $y_{ik} = \begin{cases} 1 & \text{if node } i \text{ is served by vehicle k} \\ 0 & \text{otherwise} \end{cases}$
- $I_k = \begin{cases} 1 & \text{if vehicle } k \text{ is used} \\ 0 & \text{otherwise} \end{cases}$

$$\min \sum_{e \in E} c_e \sum_{k=1}^{m} x_{ek} + \sum_{k=1}^{m} f_k I_k \tag{55}$$

$$\sum_{k=1}^{K} y_{ik} = 1 \quad \forall i \in V \setminus \{0\} \tag{56}$$

$$\sum_{k=1}^{K} y_{0k} \leq K \tag{57}$$

$$\sum_{e \in \delta(i)} x_{ek} = 2y_{ik} \quad \forall i \in V, \quad k = 1, ..., K \tag{58}$$

$$\sum_{i \in V} d_i y_{ik} \leq Z_k I_k \quad \forall k = 1, ...K \tag{59}$$

$$\sum_{e \in \delta(S)} x_{ek} \geq 2y_{hk} \quad \forall S \subseteq V \setminus \{0\}, h \in S, k = 1, ..., K \tag{60}$$

$$\sum_{e \in \delta(S)} x_{ek} \leq |S| - 1 \quad \forall S \subseteq V \setminus \{0\}, |S| \geq 2, k = 1, ..., K \tag{61}$$

$$I_k \in \{0, 1\} \quad \forall k = 1, ...K \tag{62}$$

$$x_{ek} \in \{0, 1\} \quad \forall e \notin \delta(0), k = 1, ..., K \tag{63}$$

$$y_{ik} \in \{0, 1\} \quad \forall i \in V, k = 1, ...K \tag{64}$$

$$x_{ek} \in \{0, 1, 2\} \quad \forall e \in \delta(0), k = 1, ...K \tag{65}$$

Here, constraint 56 and 57 ensure that each source node is visited exactly once and the depot is visited by no more than the number of available vehicles. Constraint 58 guarantees that the resulting graph is a set of tours. The capacity constraint 59 ensures that a source node is only assigned to a vehicle if that vehicle is used and the total amount of the source nodes assigned to a vehicle does not exceed its capacity. And constraint 61 is a subtour elimination constraint.

In previous sections, we have shown that the structure of f_k plays an important role in the complexity of the FLSPN problems. Here, $f_k(Z_k)$ in MDPP is an input parameter to the problem since Z_k and f_k are both predefined. Therefore, given the problem formulation above, how f_k is calculated does not have an impact on the complexity of the problem. However, from a capacity design perspective, if the variable cost of a MPU is linear with respect to the amount of flow it processes as opposed to its capacity, then the vehicle cost f_k term can be dropped from the objective function 55. The problem becomes a basic capacitated vehicle routing problem.

4.4.2 Time constraints of static and continuous sources

Notice that the capacity of a MPU, denoted by Z, is in terms of processing rate, e.g. tons per day. Here, we consider two types of source nodes: continuous and static. A continuous source node produces a continuous flow with no interruption in time, such as algae-based biofuel system or waste water sources. A static source node produces a fixed amount of material (biomass) only once or periodically with very long intervals, i.e. the processing and traveling time is so much shorter than the interval that the decision problem is focused on one period only. For example, terrestrial biomass is harvested only once in a year or in a season, in which case the MPUs have to process this biomass over a finite period of time post-harvest to prevent deterioration. In this case, the the periodicity of the harvest and

the periodicity of the vehicle tour are completely decoupled and the demand at a source node can be considered as fixed in one harvesting cycle. The differentiation of the two cases is not necessary for facility location problems discussed in previous sections, because the material flow from source nodes is either transported continuously through pipeline (FLSPN problems) or through direct source-sink truck trips. For a continuous pipe flow, the source nodes can be regarded as being served immediately by the facility, therefore time constraint does not exist. When direct truck trips from source nodes to the facilities are used, the frequency at which the material arrives at the facility can be controlled after facility capacity sizing. For mobile distributed processing, however, the two cases enforce different constraints on the relationship between MPU capacity and time.

Given one tour that starts and ends at the depot, let t^P be the total time a MPU spends processing flow on its tour, t^T be the total time it spends traveling on edges of this tour and t^S be the total stopping time needed at the sources that may include setup, stand-down and cleaning. Since the capacity of the MPU, Z, is in terms of processing rate (e.g. tons/day), the total amount the MPU can process in one trip is then $Z * t^P$. The total time from the MPU leaves the depot to the MPU returns to the depot is $t^P + t^T + t^S$.

The demand of a static source node is fixed, e.g. tons. If at the time of production, all biomass at the sources is harvested and available for processing, then obviously the MPU should process all biomass at a source node upon arrival before moving to the next source node. To differentiate from the continuous source nodes, we denote the total demand of all static source nodes on a tour by $\bar{D}$. The amount processed by the MPU on a given tour should be equal to the total demand on the tour, i.e. $Z * t^P = \bar{D}$. In this case, the faster that a process unit can operate, i.e. the bigger the capacity of the unit, the more stops it can make in a given harvest season. The harvested biomass needs to be processed within a certain time window post harvest, and therefore there is a constraint on the total time a MPU spends on a given tour, i.e. $t^P + t^T + t^S \leq t^{max}$. Therefore, we have the following constraint on the capacity of the unit:

$$Z \geq \frac{\bar{D}}{t^{max} - t^T - t^S} \tag{66}$$

And the above constraint enforces the following constraint on the optimization problem to replace constraint 59

$$I_k Z_k \geq \frac{\sum_{i \in V} d_i y_{ik}}{t^{max} - t^{travel} \sum_{e \in E} l_e x_{ek} - t^{stop} \sum_{i \in V} y_{ik}} \quad \forall k = 1, ..., K \tag{67}$$

where t^{travel} is the travel time of a MPU on a unit distance and t^{stop} is the stopping time required at each source.

The demand of a continuous source, on the other hand, is in terms of flow rate, e.g. tons per day. For each source node on the tour, the time between visits by the MPU is $t = t^P + t^T + t^S + \delta$, where $\delta \geq 0$ is any time the MPU spends stopping, unloading, etc. δ is important for continuous sources, because the MPU trips are continuous instead of happening once in every harvesting season. Let D be the total demand of the source nodes on the tour, the amount of flow accumulated at all the source nodes on the tour during one single trip of the MPU is $D * t$. In steady state, the amount processed by the MPU in one trip should equal to the amount accumulated during this time period, i.e. $Z * t^P = D * t \geq D * (t^P + t^T + t^S + \delta)$. If δ represents all the stopping time the MPU takes, then the above constraint is binding. We enforce the following constraint on the capacity of the MPU:

$$Z \geq (1 + \frac{t^T + t^S + \delta}{t^P}) * D \tag{68}$$

This indicates that the following constraint needs to replace constraint 59 in the optimization problem defined by 55 - 65:

$$I_k Z_k \geq (1 + \frac{\delta}{t^P} + \frac{t^{stop}}{t^P} \sum_{i \in V} y_{ik} + \frac{t^{travel}}{t^P} \sum_{e \in E} l_e x_{ek}) * (\sum_{i \in V} d_i y_{ik}) \quad \forall k = 1, ..., K \tag{69}$$

Notice that the difference between constraint 69 and 59 is because of the supply and demand being flow rate instead of fixed amount. Constraint 69 ensures that the processing rate of the MPU on a given tour is higher than the rate of flow generation of all source nodes in this tour, after consideration of the travel time and any delay in processing.

In the case of static sources, once the vehicle leaves the depot, it processes all the biomass on its tour before returning to the depot. The same trip is repeated every harvesting cycle, but the process time in each trip is fixed. Assume that the project life time, T^{life}, includes H discount period. And there are $n^{harvest}$ harvesting cycles in a discount period, then the transportation cost on edge e is given by $c_e = \sum_{h=1}^{H} \eta_h n^{harvest} c^{transport} l_e$, where $c^{transport}$ is the unit transportation cost and η_h is the discount factor for the hth discount period. For continuous sources, however, only travel time t^T and stopping time δ are fixed. The processing time on each trip, t^P, can be considered as a decision to be made by the planner prior to the routing problem. In steady state, the processing time a MPU spends at each continuous source node should be proportional to its flow rate. The planner might instead enforces a fixed total time the MPU spends on any tour. The shorter t is, the shorter the total time between visits is, and therefore the less storage space is needed at the source nodes. On the other hand, the shorter t is, the frequency of travel increases in any given period. The number of times the MPU has to travel on a given tour in the project life time is $\frac{T^{life}}{t}$. Assume that the number of trips a MPU has to make in a discount period is n^{trip}. The transportation cost on edge e is $c_e = \sum_{h=1}^{H} \eta_h n^{trip} c^{transport} l_e$. The total time between visits is determined by the tradeoff between transportation costs and storage costs. In this analysis, we consider t to be an exogenous factor that is given as input to the optimization model. It is important to notice that in this case, given a vehicle tour, a fixed t corresponds to a fixed amount of flow to be processed on each trip, which forces the problem to be the same as the static source case.

4.4.3 Algorithms and analysis for mobile distributed processing

There are many heuristics available for the vehicle routing problems. Classical heuristics developed between 1960 and 1990 produce good quality solutions within modest computing times by performing a relatively limited exploration of the search space. These heuristics can be extended to account for the diversity of the constraints in real-life and are still widely used in commercial packages [150]. In recent years, significant development has been made in metaheuristics, which usually use sophisticated neighborhood search rules and memory structures, and emphasize on performing in depth exploration of the most expected regions of the solution space. These metaheuristics produce much higher solution quality at the price of computing time.

Given the nature of the vehicle cost, it seems more nature to use a route-first, cluster-second approach [15], which finds a giant TSP tour on all the nodes first and then split it into feasible tours. As mentioned earlier, because of our focus on capacity design, any limit on the number of vehicles available of any type is artificially placed. In other words, the problem we are considering can be regarded as having unlimited number of vehicles. In this case, selecting the right vehicle type to route is the most important factor. The route-first, cluster-second algorithm below attempts to find the right combination of the fleet and their routes.

Algorithm Route First-Cluster Second

Step 1 Find an α-optimal TSP tour on all the nodes;

Step 2 Arbitrarily assign a direction to the giant tour and let 1 be the first source node on the directed tour after the depot (which we denote by 0), 2 be the second source node on the tour after the depot, ..., n be the last source node on the tour after the depot.

Step 3 Let c_{ij} be the distance matrix and define a matrix v_{ij} as the cost of the cheapest vehicle traveling to serve the source nodes $(i+1, i+2, ..., j)$ in that order if the vehicle route $(0, i+1, i+2, ..., j, 0)$ is feasible $(i < j)$ and $+\infty$ otherwise.

$$v_{ij} = \begin{cases} c_{0(i+1)} + \sum_{s=i+1}^{j-1} c_{s(s+1)} + c_{j0} + \min\{f_k, k = 1, ..., K\} & \text{if feasible} \\ +\infty & \text{otherwise} \end{cases} \quad (70)$$

v_{ij} is defined by the following algorithm:

Step 3 Find the least cost path from 0 to n in the directed graph with arc cost (v_{ij}) we will have an optimal partition of the giant tour into feasible vehicle routes.

Claim 4 *If $f_k(Z_k) = a_1 + a_2(Z_k)^s$, where a_1 and a_2 are positive and $s \in (0, 1)$, then v_{ij} in step 2 of algorithm MPUP can be found by the following algorithm for the case of static sources.*

Given a vehicle tour $Tour_{ij}$ defined by $(0, i+1, i+2, ..., j, 0)$, calculate the following constants:

- $t^T = t^{travel}(l_{0(i+1)} + \sum_{s=i+1}^{j-1} l_{s(s+1)} + l_{j0})$
- $t^S = t^{stop}(j - i)$
- $\bar{D} = \sum_{s=i+1}^{j-1} d_s$ (or $D = \sum_{s=i+1}^{j-1} d_s$ for continuous sources)

if all vehicle have been checked and are all infeasible **then**

$v_{ij} = +\infty$

else

Select the next smallest vehicle k and check the feasibility constraint: whether $\frac{\bar{D}}{Z_k} + t^T + t^S \leq t^{max}$ (or whether $\frac{Z_k}{D} \geq 1 + \frac{t^T + t^S + \delta}{t^P}$ for continuous sources)

if the current vehicle k is feasible **then**

$$v_{ij} = c_{0(i+1)} + \sum_{s=i+1}^{j-1} c_{s(s+1)} + c_{j0} + f_k \quad (71)$$

else

Go to the next smallest vehicle

end if

end if

Proof 4 *$f_k(Z_k)$ is defined by equation 44 and therefore is a monotonically increasing function of Z_k when $a_2 \neq 0$. For a given vehicle tour $(0, i+1, i+2, ..., j, 0)$, the travel cost is fixed. The minimum v_{ij} is determined by the vehicle that has the lowest vehicle cost, which is, based on function $f_k(Z_k)$, the vehicle with the smallest possible capacity that can process the demand in time.*

The route first-cluster second algorithm obviously depends on the initial quality of the TSP tour. If Christofides' algorithm is used to obtain the initial TSP tour, $\alpha = 1.5$, $O(n^3)$ [33]. The second-phase cluster problem is a standard shortest-path problem on an acyclic graph that can be solved in $O(n^2)$ by Dijkstra's algorithm [79]. Haimovich and Rinnooy Kan showed that if all customers have unit demand, this algorithm is asymptotically optimal but not for general demands except in trivial cases [18].

The subalgorithm presented above requires first sorting all the vehicle costs $O(KlogK)$, then checking the feasibility constraint for every pair of (i, j) and checking the feasibility constraints, $O(Kn^2)$. This modification to the original algorithm does not change the worse case performance of the algorithm, but for a large number of vehicle types, it is expected to improve the speed of the algorithm. The complexity of the complete algorithm proposed in this section is $O(n^3 + Kn^2 + KlogK)$.

4.5 Case Study

Two cases of study are used here to demonstrate the algorithms proposed for mobile distributed processing.

4.5.1 Case Study: Static Sources

This case study is designed to determine the optimal selection of mobile processing units and their routes. We consider one hundred locations with different unsplit feedstock supply, which are located 50 to 500 miles away from the depot. The biomass quantity available at these sources ranges from 100 to 2000 tons. For capacity design, we assume there are an infinite number of 200 types of mobile processing units, the capacity of which range from 1 to 200 tons per month. The cost of the MPUs is assumed to scale exponentially with an exponent of 0.6. For a mobile processor with capacity one ton per month, the capital cost is 8 million dollars. The transportation cost is assumed to be on average 200 dollars per mile, The project lifetime is assumed to be 20 years, and the transportation and processing is assumed to occur once per year. Therefore, with 10% annual effective interest rate, the net present value of the transportation cost is about 2000 dollars per mile. The biomass has to be processed within in a 12-month time window. The parameters used in this case study are summarized in table 6.

Table 6: Parameters used in case study for static sources

Number of biomass sources	100
Range of demand at sources	100 to 2000 tons
Range of distance from the depot	50 to 500 miles
Number of mobile processor types	200
Ranges of processors capacity	1 to 200 ton/month
Cost of MPU with 1 ton/month capacity	8 M$
Cost scale exponent	0.6
Transportation cost over project lifetime	2000 $/mile
Travel speed	10 miles/hour
Setup and stand down time at each source	1 day
Biomass must be processed within	2 months
Harvesting Cycle	1 year

The problem is to decide which type of MPUs to choose and how many to purchase. The result shows that six MPUs are needed with capacity of, respectively, 159, 180, 191,

152, 188, 175 tons/month. It is important to notice that the values used in this case study are based on best guess or order-of-magnitude estimates, the demand of distance matrices are generated randomly within the given range. This should not be considered empirical. For any practical application, more accurate data should be used given the best information available.

The optimal fleet size and routing changes when the time limit for processing changes. Table 7 shows the results when the time limit varies from one month to one year. The results shown here include the optimal number of mobile processors, average capacity of the chosen mobile processors, the average length of the tours and the total cost for the chosen mobile processors. As expected, mobile processors with relatively larger size are preferred because of economies of scale in production. The average processor capacity ranges from 100 to 200 tons per month. When time limit is short, more MPUs are needed to satisfy the time constraint with relatively shorter average tour length. When the time limit is 12 months, a big MPU with a capacity of 187 ton/month can process all the sources while meeting the time limit and it travels on the initial TSP tour determined by the algorithm. As the upper time limit increases, there are more cost reduction potentials and therefore the objective value, i.e. the total cost, decreases.

4.5.2 Case Study: Continuous Sources

Our second case study considers 150 sources that generate chemical flow on a continuous basis, which are located 50 to 500 miles away from the depot. The flow rate at these sources ranges from 100 to 10,000 gallons per day. For capacity design, we assume there are 1000 types of mobile processing units, the capacity of which range from 0.001 to 1 million gallons per day. The cost of the MPUs is assumed to scale exponentially with an exponent of 0.6. For a mobile processor with a capacity of one hundred gallons per year, the capital cost is 1 million dollars. The transportation cost parameters are the same with static sources. In the baseline case, we assume that the maximum time a MDP is allowed

Table 7: Results of optimal number of MPUs and their average size fluctuating with the time limit in processing the biomass

Time limit	No. of MPUs	Average MPU capacity	Average tour length	Objective Value
1 month	13	174 ton/month	1347 miles	2332 M$
2 months	6	190 ton/month	2918 miles	1154 M$
3 months	4	189 ton/month	4378 miles	778 M$
4 months	3	188 ton/month	5837 miles	590 M$
5 months	3	137 ton/month	5837 miles	494 M$
6 months	2	187 ton/month	8755 miles	404 M$
7 months	2	150 ton/month	8755 miles	358 M$
8 months	2	125 ton/month	8755 miles	325 M$
9 months	2	108 ton/month	8755 miles	300 M$
10 months	2	110 ton/month	8755 miles	277 M$
11 months	2	103 ton/month	8755 miles	250 M$
12 months	1	187 ton/month	17510 miles	220 M$

on a single tour is 30 days. The parameters used in this case study are summarized in table 8.

Table 8: Parameters used in case study for continuous sources

Number of biomass sources	150
Range of demand at sources	100 to 10,000 gal/day
Range of distance from the depot	50 to 500 miles
Number of mobile processor types	1000
Ranges of processors capacity	0.001 to 1 Mgal/day
Cost of MPU with 1 MGY processing ratey	0.01 M$
Cost scale exponent	0.6
Unit distance transportation cost	100 $/mile
Travel speed	10 miles/hour
Setup and stand down time at each source	6 hours
Stopping time at the depot	2 days
Process time limit of each trip	30 days
Discount Period	1 year

For continuous sources, the result changes with the frequency that the planner chose for the vehicle tours. In this case study, we take as input the maximum time a mobile processor can take on one single tour. And table 9 shows the result when this maximum time varies

from 10 days to 100 day.

Table 9: Results of optimal number of MPUs and their average size fluctuating with the single tour duration

Single tour duration	No. of MPUs	Average MPU capacity	Average tour length	Objective Value
10 days	8	350,000 gal/day	2032 miles	584 M$
20 days	4	488,000 gal/day	4063 miles	292 M$
30 days	2	931,000 gal/day	8127 miles	195 M$
40 days	2	545,000 gal/day	8217 miles	146 M$
50 days	2	434,000 gal/day	8127 miles	117 M$
60 days	2	512,000 gal/day	8127 miles	98 M$
70 days	2	437,000 gal/day	8127 miles	84 M$
80 days	2	486,000 gal/day	8217 miles	73 M$
90 days	1	893,000 gal/day	16253 miles	65 M$
100 days	1	822,000 gal/day	16253 miles	58 M$

Table 9 shows the results when the travel frequency predetermined by the planner varies from 10 days to 100 days. In this case, the number of mobile processors does not vary significantly. The capacities of mobile processors chosen by the algorithm are the larger ones within the given 0.001 to 1 million gallons per day range. The total capacity of all MPUs, which is calculated by the average MPUs times the number of the MPUs, decreases as the single tour duration increases. This is because the more frequent the visits, the more time the MPUs have to spend on stopping and travel and therefore the more capacity is required to be able to finish processing all the flows in the same time period. Subsequently, the higher the visiting frequency is, the higher the total cost is.

4.6 Conclusion and Future Research

In this analysis, we studied two types of transportation problems in the biofuel industry: facility location with an underlying pipeline network and the routing problem of mobile distributed processors. This analysis aims at understanding the impact of concave production cost on routing and location decisions, independent of other complicating factors in

the production planning problems. For facility location and sizing with pipeline network with general facility cost, we have proved that when no fixed facility cost exists, a linear facility cost can be neglected and we have provided a polynomial algorithm to solve the problem to optimality. When a fixed facility cost exists, the problem immediately becomes NP-hard. And in fact, the concavity in the facility cost does not make the problem harder. We provided a polynomial algorithm that produces a good solution to these two cases. An overview of facility location problems on bipartite graph and vehicle routing problem was provided. Several variants of the problem covered in the the existing literature have been reviewed and mathematical formulations are provided. For mobile distributed processing, we formulated the problem as a capacitated vehicle routing problem with time constraints. The route-first-cluster-second algorithm is modified based on the nature of cost concavity. Static and continuous sources are differentiated in the formulation and algorithm with different time constraints and considerations. We presented a static source case study with one hundred static sources and 200 types of mobile processors and a continuous source case study with 150 sources and 1000 types of mobile processors.

Formulations and algorithms in this paper are designed for general situations. Other important issues may arise in practice. For example, another type of economy of scale may arise when the cost of producing a large amount of production units of the same capacity is cheaper than producing the same amount of production units of a variety of capacities. This type of cost scaling, though not very common in the chemical processing industry, is common for combine harvesters and machinery in the agriculture industry. The other interesting field for future research is the stochastic and dynamic components of the problem that involve uncertainty, such as uncertainty in the supply at each source node due to interruption or break-downs.

Chapter V

CONCLUSION AND FUTURE RESEARCH

5.1 Summary

This thesis provides solution methods to the design of spatially distributed large-scale biofuel systems, with a particular focus on bioreactor-based fuel production. Bioreactor systems are common in chemical industry, but production system with bioreactors that are highly distributed over a wide geographic area is an evolving field because of recent technology advancement in fuel production from micro-algae. Such bioreactor systems for biofuel production are expected to be of very large scale due to increasing demand of renewable fuel. The four key problems from the production planner's point of view are 1) how to choose the most economical and most environmental friendly process design; 2) how to determine the capacity of the production equipment to build; 3) where should the production equipment be positioned; and 4) how to transport the material flow from the sources to the production sites. This thesis integrates tools from different fields to develop prescriptive models that answers these questions in production systems design.

Chapter 2 introduces a new technology for producing ethanol from blue-green algae. Algae are grown in closed photobioreactors and generate ethanol-water solution, which is then transported to a series of separation processes to be converted into fuel grade ethanol. This new technology forms the case study which is used throughout the thesis. Chapter 2 models the energy efficiency and greenhouse gas emissions of three different type of process designs by integrating lifecycle assessment with chemical process modeling. Vapor compression steam stripping (VCSS) combined with membrane separation is shown to have the lowest emission profile and the lowest energy consumption. An important factor introduced in Chapter 2 is the initial ethanol concentration in the ethanol-water stream

coming out of the bioreactors. All energy and emission level are evaluated under a range of possible values of this initial concentration.

Following Chapter 2, Chapter 3 studies the optimal production capacity design for bioreactor-based fuel systems, and the algae-based ethanol technology described in Chapter 2 is used as a case study. Two nonlinear models based on net present value maximization are presented for an integrated process design and a distributed-centralized design with respectively, 18 and 20 parameters. Energy cost as a function of the initial concentration, which was derived in Chapter 2, is incorporated into the capacity design models in Chapter 3. The optimal basic production capacity for a reference case is shown to be 320 million gallons ethanol per year, which is similar to the size of an optimal cellulosic ethanol plant. Global sensitivity analysis based on Sobol's indices shows that 5 parameters in the integrated VCSS and membrane design are the major contributor to the overall variation in the optimal capacity decision and the corresponding net present value: ethanol price, bioreactor level flow rate, bioreactor building speed, initial ethanol concentration, transportation scaling exponent, project lifetime and the discount factor. Conditional Value at Risk optimization approach is used to evaluate the optimal production capacity when the planner is risk averse. Based on the models presented in Chapter 3, the planner, based on the best information available, could study how exogenous and indigenous factors can influence his optimal choice and how his risk preference may influence this decision.

While Chapter 3 focuses on the influence of detailed engineering parameters on production capacity design, the analysis is based on completely symmetric network formed by equivalent sized bioreactors. Many bioreactor-based fuel systems are, however, not symmetric in nature. Motivated by this fact, Chapter 4 studies the production equipment location, capacity sizing decisions as well as transportation routing decisions when the location and demand of the sources (e.g. bioreactors) are random and given. The problems studied in Chapter 4 are not restricted to bioreactor-based systems, but are also applicable to terrestrial biomass and other production systems that involve flow collection from

spatially distributed sources. Several fixed and mobile processing options in the biofuel industry are formulated as network flow and capacitated vehicle routing problems. Polynomial algorithms are provided that allows sensitivity analysis of the parameters involved. To demonstrate the algorithms for mobile distributed processing, we provide two case studies that involves, respectively, biofuel production from terrestrial biomass and biofuel production from continuous liquid sources.

5.2 Future Directions

Chapter 2 and Chapter 3 studied the bioreactor-based fuel production system from both environmental and economical perspective. These two aspects, however, were not completely integrated. As explained in the conclusion remark of Chapter 3, emission factors and carbon price can be integrated into the capacity decision model. The influence of this parameter on the optimal capacity decision could be interesting to the planner, considering that carbon price is an unavoidable topic for biofuel production.

Chapter 3 has shown that the cost scaling effect of transportation cost is an important factor in the production capacity design of bioreactor-based fuel systems. This scaling factor is derived in Appendix B based on several predefined geometry. However, piping cost, especially the operation cost of piping, depends on both pipeline length and diameters. The optimum pipe diameters are the ones that provide the minimum total cost of both operational power and fixed charges for the particular piping system [114]. Many studies have been done in pipeline network design: Tondeur and Kvaalen [147] proposed that the optimal configuration of transfer or separation processes is one which satisfies or approaches equipartition (uniform distribution) of quantities related to entropy production. Singh and Mahar [140] developed an optimal design method for a multidiameter, multioutlet pipeline employing a dynamic programming approach. Optimal lengths of varying pipe sizes for a multioutlet pipeline were obtained by minimizing the annualized cost of the pipeline, which can include either fixed costs, energy costs or both [148]. Wechsatol et al. [160]

investigated several classes of flow systematically in a T-shaped construct with fixed internal and external size to show the effect of junction losses on the optimized geometry of tree-shaped flows. Yang and Ogden [175] developed models to characterize delivery distances and to estimate costs, emissions and energy use from various parts of the hydrogen delivery chain including pipeline delivery. However, the literature on engineering design of pipeline networks is disconnected from the economic analysis literature. Little research has been done on the effect of pipeline network design on the production capacity decision. Therefore, analysis on the cost scaling of piping with considerations of detailed pipeline engineering design is an interesting and important field of research.

Chapter 4 provided several fundamental deterministic problems in the biofuel industry. The aim of this Chapter is to understand the impact of concave production cost on routing and location decisions, independent of other complicating factors in the production planning problems. However, many interesting properties in the biofuel industry can be important extensions to the work presented in Chapter 4. For example, another type of economy of scale may arise when the cost of producing a large amount of production units of the same capacity is cheaper than producing the same amount of production units of a variety of capacities. This type of cost scaling, though not very common in the chemical processing industry, is common for combine harvesters and machinery in the agriculture industry. Chapter 4 explained three different underlying transportation problems: facility location on bipartite graph, facility location on complete graph and vehicle routing problem. These three problems may all be potential choices for producing the same final product, e.g. ethanol, depending on different process designs. A comparative study on how the total cost behaves under different process designs and the potential for cost reduction is an important field of future study. And finally, many stochastic and dynamic components of the problems in the biofuel industry involve uncertainty, such as uncertainty in the supply at each source node due to interruption or break-downs. Adding stochasticity to the formulation in Chapter 4 is an interesting future research field.

Appendix A

SUPPORTING INFORMATION FOR CHAPTER 2

Method: Weight Percentage of Ethanol Achieved with Vapor Compression Steam-Stripping.

Given the concentration of the input stream from the photobioreactor to the vapor compression steam stripping column, the output concentration of the steam stripping column can be calculated based on thermodynamics.

Vapor pressure

The Antoine equation is used to calculate the saturated vapor pressure of the streams in an ideal mixture.

$$P_i^o = 10^{A_i - \frac{B_i}{C_i + T}} \tag{72}$$

where T is the operating temperature of the column, and A, B, C are the component-specific constants. Table 10 shows the component-specific constants for ethanol and water used in this calculation for T in $^\circ C$.

Partial vapor pressure

The partial vapor pressure in the column is calculated as:

$$P_i = P_i^o \times x_i \times \gamma_i \tag{73}$$

Table 10: Antoine equation coefficients and temperatures

A	A	B	C	T_{min}	T_{max}
Water	8.0713	1730.63	233.426	1	100
Ethanol	7.5867	1281.59	193.768	78	203

where x_i is the mole fraction of component i in the input stream; γ_i is the activity coefficient of component i ; and P_i^o is the ideal behavior of component i .

For the calculation of activity coefficient, we apply van Laar equations, as follows:

$$\ln \gamma_1 = A_{12}(\frac{A_{21}x_2}{A_{12}x_1 + A_{21}x_2})^2 \tag{74}$$

$$\ln \gamma_2 = A_{21}(\frac{A_{12}x_2}{A_{21}x_1 + A_{12}x_2})^2 \tag{75}$$

where A_{12}, A_{21} are binary interaction parameters; and x_1, x_2 are mole fractions. For an ethanol (1)-water (2) mix, the values of the parameters are $A_{12} = 1.6798$, $A_{21} = 0.9227$. The mole fraction of component *i* in the output stream is, therefore, calculated as: $R_i = \frac{P_i}{\sum_i P_i}$.

Vapor Compression Steam Stripping (VCSS)

For an input concentration of ethanol into the VCSS of 1 wt%, the mole fractions of ethanol and water are 0.394% and 99.61% respectively. The operating temperature is 93.98 $^\circ C$. The ideal behavior of the components is as follows.

$$P^o_{water} = 10^{8.07131-\frac{1730.63}{233.426+98.98}} \times 0.001316 \quad atm/torr = 0.9661 \quad atm \tag{76}$$

$$P^o_{EtOH} = 10^{7.5867-\frac{1281.59}{193.768+98.98}} \times 0.001316 \quad atm/torr = 2.1289 \quad atm \tag{77}$$

Therefore, the partial pressure of the components in the column is:

$$P_{water} = P^o_{water} \times x_{water} \times \gamma_{water} = 0.9961 \quad atm \times 99.61\% \times 1.0000 = 0.9623 \quad atm \tag{78}$$

$$P_{EtOH} = P^o_{EtOH} \times x_{EtOH} \times \gamma_{EtOH} = 2.1289 \quad atm \times 0.394\% \times 5.2376 = 0.04393 \quad atm \tag{79}$$

The mole fraction of ethanol in the output stream of the VCSS is, then:

$$R_i = \frac{P_{EtOH}}{P_{EtOH} + P_{water}} = \frac{0.03723}{0.7998 + 0.03723} = 4.36\% \tag{80}$$

Assuming 90% distillation efficiency, the weight percentage of ethanol in the output stream is

$$\frac{4.36\% \times 46}{4.36\% \times 46 + (1 - 4.36\%) \times 18} \times 90\% = 9.4\% \tag{81}$$

Method: Evaporation Energy

During the vapor compression steam stripping processes, the total heat needed is calculated based on the total heat of evaporation. A condenser is used for heat recovery, i.e. the energy of evaporation is captured and recycled through the condenser. The energy of the condenser in this process is given by: $E_{condensor} = \sum_i m_i \Delta_v H(T)_i$, where m_i is the mass flow rate of chemical i , and is the heat of evaporation of chemical i .

Therefore, the following equation is used for calculating the total heat needed in the distillation process:

$$E_{evaporation} = \frac{m_{EtOH}\Delta_v H(T)_{EtOH} + m_{water}\Delta_v H(T)_{water}}{m_{EtOH} \times (26.74 \quad MJ/kg_{EtOH})} \tag{82}$$

Throughout the calculation, 838 kJ/kg is used as the heat of evaporation of pure ethanol $\Delta_v H(T)_{EtOH}$, and 2260 kJ/kg is used as the heat of evaporation of water $\Delta_v H(T)_{water}$. Based on the equation 82, the VCSS total heat is:

$$E_{evap} = \frac{0.0941 \times (838 \quad kJ/kg) + (2260 \quad kJ/kg) \times (1 - 0.0941)}{0.0941 \times (26.74 \quad MJ/kg)} \approx 0.845 \quad MJ/MJ_{EtOH} \tag{83}$$

For 80% heat exchange efficiency, the external heat needed is 20% of the above number, 0.169 MJ/MJ_{EtOH}. For 90% heat exchange efficiency, the external heat needed is 10% of the above number, 0.0845 MJ/MJ_{EtOH}.

Method: Compression Energy

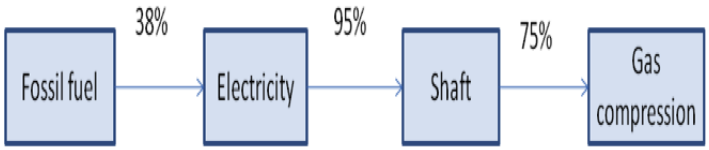


Figure 20: Compression Energy Supply Chain (38% thermal efficiency of electricity production is illustrative only)

A compressor is used in the VCSS process. An upper bound on the work required is the adiabatic work. Under adiabatic conditions, $PV^g = C$ and under an open flowing system the work can be expressed as

$$W = \int_{V_{in}}^{V_{out}} VdP = \frac{\gamma nRT}{\gamma - 1}[(\frac{P_{out}}{P_{in}})^{\frac{\gamma-1}{\gamma}} - 1] \tag{84}$$

where n is the molar flow of the stream, T is the operating temperature in Kelvin; and P_{out}, P_{in} are the input pressure and output pressure respectively; and R is the gas constant, 8314 $J/kmol/K$, and γ is the adiabatic coefficient. Note that for compression via volume reduction, not used here, the expression for adiabatic work is somewhat different:

$$W = \int_{V_1}^{V_2} PdV = \frac{\gamma nRT_1}{\gamma - 1}[(\frac{P_2}{P_1})^{\frac{\gamma-1}{\gamma}} - 1] \tag{85}$$

Figure 20 shows the flow of the compression energy. Electricity production is assumed to be 38% efficient. The electrical energy input drives the rotation of the shaft, during which 5% of the input energy will be lost. Finally, we assume 75% isentropic process efficiency when the movement of the shaft is translated into gas compression.

Finally, to translate the calculation into MJ of ethanol basis, the energy density of ethanol (LHV) 26.74 MJ/kg is incorporated.

The adiabatic energy needed for VCSS compression in the 1% case is

$$
\begin{aligned}
W_{adiabatic} &= \frac{1.3 \times (5.27 \quad kmol) \times (8314 \quad J/kmol/K) \times (93.98 + 273.15)K}{0.3} \\
&= \times[(\frac{101.3 \quad kPA}{81.56 \quad kPA})^{\frac{0.3}{1.3}} - 1] \times \frac{1}{26.74 \quad MJ/kg} \times \frac{1}{8.998 \quad kg} \\
&= \approx 0.0149 \quad MJ/MJ_{EtOH}
\end{aligned} \tag{86}
$$

$$
E_{comp_total} = W_{adiabatic} \times \frac{1}{75\%} \times \frac{1}{95\%} \times \frac{1}{3.6 \quad MJ/kWhe} \approx 0.0058 \quad kWhe/MJ_{EtOH} \tag{87}
$$

Fertilizers

Production of synthetic nitrogen fertilizer is reported to require an average of 23.7 MJ/kgN with greenhouse gas emissions of 1.675 $kgCO2e/kgN$; production of phosphorus fertilizer is reported to require 5.78 $MJ/kgP2O5$ with greenhouse gas emissions of 0.97 $kgCO2e/kgP2O5$ [85]. We estimate that fertilizer transportation will contribute an additional 1 MJ/kg. Correspondingly, the energy required to produce and transport the fertilizers are 0.0017 MJ/MJ_{EtOH} and 0.000017 MJ/MJ_{EtOH} for nitrogen and phosphorus respectively. The greenhouse gas emissions are 0.11 $gCO2e/MJ_{EtOH}$ and 0.0026 $gCO2e/MJ_{EtOH}$ for nitrogen and phosphorus fertilizers respectively. Nitrous oxide can be expected to be emitted in small quantities from the nitrogen that is in the bioreactor system. Nitrous oxide emissions factors are not well characterized, particularly in aquatic environments; the IPCC recommends a default value of 1% (0.3-3%) of applied nitrogen that is released as N2O-N [46]. This corresponds to 0.0157 g of N2O (0.005 to 0.047 g of N2O) for each gram of applied nitrogen, corresponding to a greenhouse gas emissions of 0.3 $gCO2e/MJ_{EtOH}$.

Bioreactor Production The energy to produce polyethylene, not including the embodied energy in the product, is reported to be 22.39 MJ/kg. Based on reported production energy use and emissions, we calculate that the greenhouse gas emissions from production of polyethylene is 1.28 kg CO2e per kg polyethylene [21] . The photobioreactors are 50 feet long with a circumference of about 12.6 feet. We estimate a wall thickness

of 0.2 mm for the purposes of these calculations. Our estimates are 0.017 MJ/MJ_{EtOH} for the production energy and 1.0 g $CO2e/MJ_{EtOH}$ greenhouse gas emission contribution. The energy and greenhouse gas emissions associated with manufacturing and installing the photobioreactors have not been calculated, but are expected to be small compared to the energy to produce the polyethylene.

Transportation Energy For general truck freight, for fertilizer transport, we use 1 $GJ/ton-mile$. For ethanol transport by truck we use 21.7 MJ per vehicle mile [40], and an assumed tanker capacity of 25 tons to derive 0.9 $GJ/ton-mile$ for the energy intensity of ethanol transportation. For rail transport we use 0.36 $GJ/ton-mile$ and for barge freight we use 0.54 $GJ/ton-mile$ [40]. These values are consist with the calculations of the petroleum baseline assessment developed by Skone et al [142]. We use lifecycle emissions of 90 $gCO2e/MJ$ of diesel transportation fuel [142].

Site Preparation The photobioreactor site is assumed to be fairly level land. Grading and other site preparation activities will require roughly 25 liters of diesel fuel per hectare [72]. Assuming a site lifetime of 20 years, site preparation will contribute roughly 0.00025 MJ/MJ_{EtOH} for ethanol productivity of 56,000 liters per hectare per year.

Ethanol Distribution (Transportation) The ethanol will be transported from the production facility to the storage/blending site, and then distributed to refueling stations. Ethanol transportation distances and modes can vary significantly [158]. We assume that average US ethanol is transported by a mix of barges (40%), railroad tankers (40%), and trucks (20%), with the transported distances (one-way) of 520, 800, and 80 miles for each mode respectively [159]. The energy intensity of the transportation modes are taken to be 0.9 $MJ/ton-mile$ for truck freight, 0.36 $MJ/ton-mile$ for rail and 0.54 for barge freight. The resulting estimate of energy use in transportation is 0.017 MJ/MJ_{EtOH}.

Ethanol Combustion in Vehicle Since all of the carbon in the ethanol is from the atmosphere, or is from fossil fuel combustion and would have been emitted to the atmosphere, the emission of carbon dioxide from combustion does not add net carbon dioxide

to the atmosphere. However, a small amount of other greenhouses gas can be expected to be emitted as a result of combustion. An estimated 0.0031 g CH4 and 0.0024 g N2O are emitted per MJ from ethanol combustion [159], with a result of 0.84 $gCO2e/MJ_{EtOH}$ for the net greenhouse gas emissions from combustion of the ethanol in a vehicle.

Table 11: Energy use composition for initial ethanol concentration ranges from 0.5% to 5%

Initial concentration (wt%)	Process electric energy (kWh/MJ	VCSS heat (80%)* (MJ/MJ)	VCSS heat (90%)* (MJ/MJ)	Molecular Sieves heat (MJ/MJ)	Carbonation, pumping, sterilization mixing (kWhe/MJ)	Off-site energy (MJ/MJ)
0.5	0.01865	0.319	0.146	0.0561	0.007	3.862
0.6	0.01666	0.266	0.121	0.0561	0.007	3.862
0.7	0.01516	0.229	0.103	0.0561	0.007	3.862
0.8	0.01376	0.201	0.09	0.0561	0.007	3.862
0.9	0.01288	0.179	0.08	0.0561	0.007	3.862
1	0.01223	0.161	0.072	0.0561	0.007	3.862
1.5	0.01018	0.109	0.048	0.0561	0.007	3.862
2	0.00917	0.083	0.036	0.0561	0.007	3.862
2.5	0.00846	0.067	0.028	0.0561	0.007	3.862
3	0.00811	0.057	0.023	0.0561	0.007	3.862
3.5	0.00765	0.049	0.02	0.0561	0.007	3.862
4	0.00745	0.044	0.017	0.0561	0.007	3.862
4.5	0.00708	0.039	0.015	0.0561	0.007	3.862
5	0.00686	0.036	0.014	0.0561	0.007	3.862

* Heat exchange efficiency

**Scenario 1: Grid Electricity + Natural Gas; Scenario 2: Natural gas CHP; Scenario 3: Natural gas CH + 10 hour/day Solar

Table 12: Process-related greenhouse gas emissions for cyanobacterial ethanol production, based on 1% initial ethanol concentration and for three process energy generation scenarios (gCO_2e/MJ_{EtOH})

Stage	Scenario 1** (g CO2E/MJ)	Scenario 2 (g CO2E/MJ)	Scenario 3 (g CO2E/MJ)
Water Pumping	0.49	0.19	0.25
Condensate Pumping	0.024	0.095	0.124
Mixing	4.14	1.64	2.13
Scrubber	0.01	0.004	0.005
Sterilization	0.004	0.002	0.02
Compression Energy	8.56	3.38	4.41
Stripping and distillation (80%)	8.13	8.13	4.74
Stripping and distillation (90%)	3.64	3.64	2.12
Molecular Sieves	2.83	2.83	1.65
Total 80%/Total 90%	28.3/23.8	20.1/15.6	17.2/14.6

Table 13: Total energy use and greenhouse gas emissions for targeted scenarios with 80% heat exchange efficiency

Initial concentration (wt%)	Total energy use (S2) (MJ/MJ)	Total GHG emission (S1) (g CO2/MJ)	Total GHG emission (S2) (g CO2E/MJ)	Total GHG emission (S3) (g CO2E/MJ)
0.5	0.552	40.7	29.8	24.1
0.6	0.488	36.6	26.6	21.8
0.7	0.443	33.7	24.3	20.2
0.8	0.407	31.3	22.5	18.9
0.9	0.38	29.6	21.2	17.9
1	0.359	28.3	20.1	17.2
1.5	0.296	24.2	16.9	14.9
2	0.264	22.2	15.3	13.8
2.5	0.245	20.9	14.3	13.1
3	0.232	20.1	13.7	12.6
3.5	0.222	19.4	13.2	12.2
4	0.215	19	12.9	12
4.5	0.209	18.5	12.6	11.7
5	0.204	18.2	12.3	11.6

Table 14: Total energy use and greenhouse gas emissions for targeted scenarios with 90% heat exchange efficiency

Initial concentration (wt%)	Total energy use (S2) (MJ/MJ)	Total GHG emission (S1) (g CO2/MJ)	Total GHG emission (S2) (g CO2E/MJ)	Total GHG emission (S3) (g CO2E/MJ)
0.5	0.379	32	21.1	19
0.6	0.343	29.3	19.3	17.6
0.7	0.317	27.4	18	16.5
0.8	0.297	25.8	17	15.6
0.9	0.282	24.6	16.2	15
1	0.27	23.8	15.6	14.6
1.5	0.235	21.1	13.8	13.1
2	0.217	19.8	13	12.4
2.5	0.206	18.9	12.4	11.9
3	0.199	18.4	12	11.6
3.5	0.193	17.9	11.7	11.4
4	0.189	17.7	11.5	11.2
4.5	0.185	17.3	11.3	11
5	0.182	17.1	11.2	10.9

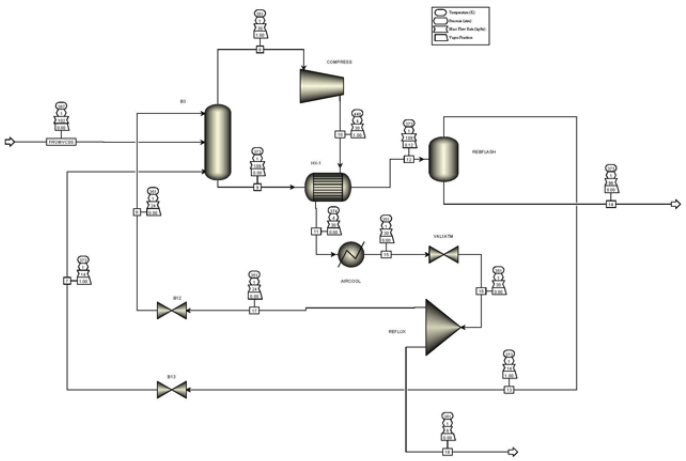


Figure 21: Ethanol concentration from 32wt% to 94wt% by distillation with vapor recompression.

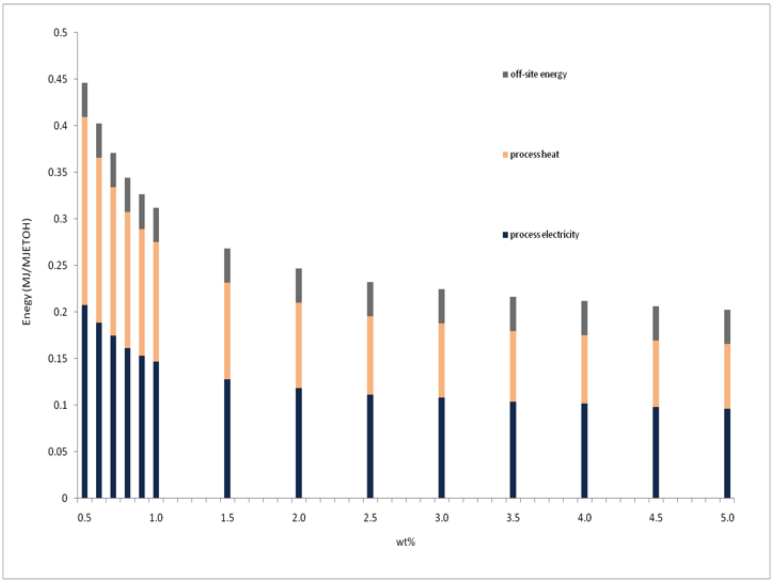


Figure 22: Lifecycle energy use assuming 90% heat exchange efficiency for cyanobacterial ethanol production using a natural-gas-fueled combined heat and power system, including process electricity, process heat, and off-site energy use.

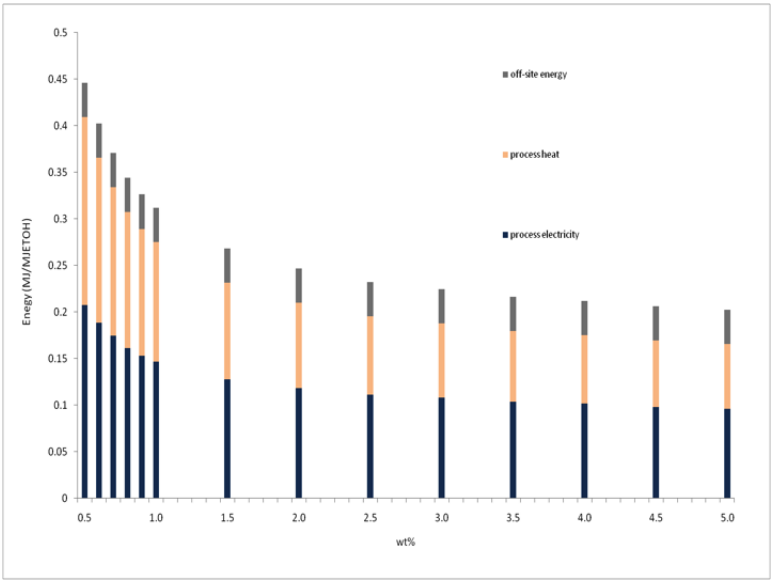


Figure 23: Lifecycle greenhouse gas emissions assuming 90% heat exchange efficiency in the case of natural-gas-fueled combined heat and power system, including the contribution from process electricity, process heat, and off-site and fugitive emissions.

Appendix B

SCALING OF TRANSPORTATION COST IN BIOREACTOR-BASED BIOFUEL PRODUCTION

B.1 Introduction

Delivery is a major cost component in most chemical processes. While the scaling effects of the purification process are similar in all chemical processes, the cost of delivery may differ. For petroleum-based chemical production, the feedstock is obtained from a single mine mouth or well head. Therefore, the feedstock delivery cost only scales linearly with the distance from the source to the refinery. In the case of corn ethanol and cellulosic ethanol, corn kernel or corn stover have to be obtained from a set of farm gates around the refinery. These farms could be distributed over a very wide geographic area depending on the size of the plant.

The cost of delivering biomass from the surrounding farmland is expected to increase with production rate due to the low density of biomass and the scarcity of resources. As a result, the optimal plant size is also dependent on the biomass yield in the region [27]. The delivery cost can be estimated based on an assumed geometry of the biomass supply region surrounding the facility [78]. Circular [110] [57] and rectangular [78] supply regions have been studied with the facility located in the center of the region. Alternatively, the delivery cost has also been assumed to follow equation 88

$$G = G_0(\frac{Q}{Q_0})^z \tag{88}$$

where Q_0 is a benchmark production rate with transportation cost G_0, and z is the dimensionless scaling exponent, which is reported to range from 1.5 to 2 [169] [109]. A detailed cost profile for harvest, purchase, and transport of biomass is provided in [87]

and [27]. Transporting biomass as sludge through pipeline instead of truck has also been studied in the literature. [86]

Chapter 3 has shown that the scaling effect of transportation cost plays an important role in the optimal production capacity decision of bioreactor-based fuel systems. Here, we derive transportation cost as a function of total production rate in such systems, in which chemical mixture has to be collected from spatially distributed bioreactors and transported to a central location for processing. The flow rate out of each bioreactor is much slower than the speed of industrial processing, which means that a reasonable scale of production requires a large number of bioreactors and therefore results in intensive liquid or gas transport. Transportation area increases with the production rate. This production layout is typical in algae-based biofuel plants and certain pharmaceutical plants.

Liquid and gas are commonly transported through pipeline networks. However, when the flow rate is very slow, intermediate storage sites can be built to store the chemical mixture and trucks can be used to transport the chemical mixture periodically. In this analysis, we first consider the case in which trucks are used to transport the flow with constant cost per ton-mile. Then, pipeline cost is calculated under the assumption that cost scales linearly with pipeline length. To explore piping cost further, we apply optimization models for pipeline diameter selection and derive capital as well as operational cost of piping as a function of production rate. Three types of pipeline routing are used for comparison. Variables are defined as follows:

- A is the total area of a production system;
- X is the length of the production system;
- Y is the width of the production system;
- x is the length of a bioreactor;
- y is the width of a bioreactor;

- n is the number of bioreactors in the production system;
- q is the flow rate out of each bioreactor;
- Q is the production rate of the production system;
- $Q_c = nq = \frac{Q}{\alpha}$ is the the total flow rate of the chemical mixture with concentration α that needs to be transported for a production system;
- G is the total cost of chemical mixture transport in the production system per period
- G_t is the cost of transporting one volume of chemical mixture by truck per distance
- G_p is the cost of transporting chemical mixture by pipeline per unit length
- r is the ratio between width and length of the production system and bioreactor, $r = \frac{Y}{X}$
- m is the number of bioreactors along one coordinate in each of the four subregion of a production system
- τ is the operating time per time period
- σ is the number of the truck visits a bioreactor per time period
- w is the average size of the trucks
- D^{T1} is the total truck transport distance of scenario T1
- D^{T2} is the total truck transport distance of scenario T2
- D^{P1} is the total pipeline transport length of scenario P1
- D^{P2} is the total pipeline transport length of scenario P2 (the dichotomic graph)
- D^{P3} is the total pipeline transport length of scenario P3, (the tetratomic graph)

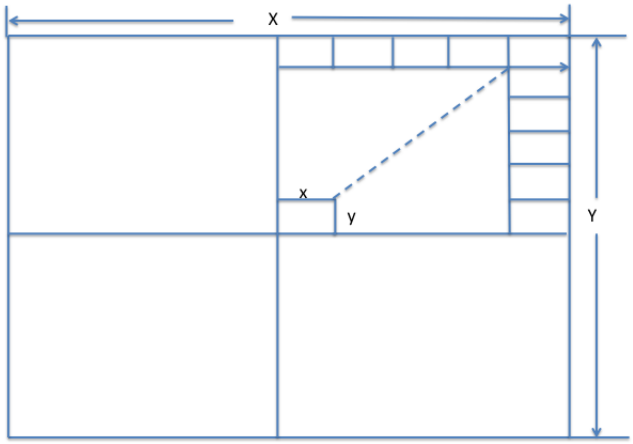


Figure 24: Transportation grid for chemical mixture supplies from a rectangular region to a central processing location

B.2 Truck transport cost

We are considering a production system which consists of one center-located production unit and spatially distributed bioreactors. An optimal production system is characterized by an optimal production rate Q^* and the corresponding land area A^*.

Consider a rectangular area $A = XY$ where X and Y are, respectively, the length and width of the production system. The production system consists of n bioreactors each of which has length x and width y. chemical mixture flow are collected from each of these bioreactors.

B.2.1 Truck Transport Scenario $T1$

In scenario $T1$, similar to the scenario for the biomass utilization facility described by Jekins [78], the chemical mixture from each bioreactor is transported separately to the center for processing. Define m as the number of bioreactors along each coordinate in one subregion in figure 24 , i.e. $m = \frac{X}{2x} = \frac{Y}{2y}$. Then $n = 4m^2$ and

$$Q = Q_c\alpha = 4m^2q\alpha \tag{89}$$

Each bioreactor $x \times y$ supplies a stream of chemical mixture. When the chemical mixture flow rate is very low, it might be economical to build intermediate storage tanks at each bioreactor and transport the chemical mixture periodically by truck. The cost of truck transport is measured in cost per volume per distance transported. The chemical mixture will flow to a storage tank at each bioreactor and be collected by the truck. We assume that the chemical mixture will be transported in batches, with size w per bioreacoter per unit time. Then,

$$\sigma w = q\tau \tag{90}$$

Notice that the dimension and size of each bioreactor $x \times y$ is fixed and does not change with production rate. Let $r = \frac{y}{x}$. If each batch of chemical mixture is transported by equation 89 and 93, we can derive the total transport distance of truck tours in one time period, D^{T1} as a function of Q,

$$\begin{aligned} D^{T1} &= 8\sigma \sum_{i=0}^{m-1}\sum_{j=0}^{m-1}(ix + jy) = 4\sigma m^2(m-1)(x+y) \\ &= 4\frac{q\tau}{w}(1+r)xm^2(m-1) \\ &= \frac{\tau(1+r)x}{2wq^{\frac{1}{2}}\alpha^{\frac{3}{2}}}Q^{\frac{3}{2}} - \frac{\tau(1+r)x}{w\alpha}Q \end{aligned} \tag{91}$$

Let G_t be the unit cost of transportation by truck ($/$m$), T be the number of time periods in project lifetime, and η be the effective discount factor per time period. The total transportation cost in the project lifetime is then,

$$G^{T1} = G_tT\eta D^{T1} = \frac{G_tT\eta\tau(1+r)x}{2wq^{\frac{1}{2}}\alpha^{\frac{3}{2}}}Q^{\frac{3}{2}} - \frac{G_tT\eta\tau(1+r)x}{w\alpha}Q \tag{92}$$

B.2.2 Truck Transport Scenario $T2$

Alternatively, a truck can pick up the chemical mixture along a horizontal tour from the center of the rectangular region to the end of a x coordinate. In this case,

$$\sigma w = mq\tau \tag{93}$$

$$\begin{aligned} D^{T2} &= 8\sigma \sum_{j=0}^{m-1} (\frac{X}{2} + jy) = 4\sigma(2m^2 x + m(m-1)y) \\ &= 4\frac{q\tau}{w} x[(2+r)m^3 - rm^2] \\ &= \frac{\tau x(2+r)}{2wq^{\frac{1}{2}}\alpha^{\frac{3}{2}}} Q^{\frac{3}{2}} - \frac{\tau x r}{w\alpha} Q \end{aligned} \tag{94}$$

The total transportation cost in the project lifetime is

$$G^{T2} = G_t T \eta D^{T2} = \frac{G_t T \eta \tau x(2+r)}{2wq^{\frac{1}{2}}\alpha^{\frac{3}{2}}} Q^{\frac{3}{2}} - \frac{G_t T \eta \tau x r}{w\alpha} Q \tag{95}$$

B.2.3 Role in capacity sizing

The scaling effect of transportation cost is an important factor in determining the size of a production system. Typically, processing cost is assumed to have economies of scale as follows

$$C = C_b (\frac{Q}{Q_b})^s \tag{96}$$

where s is less than one. Therefore, if transportation scales superlinearly with production rate, an optimal production capacity can be derived from the tradeoff between transportation cost and processing cost.

In the case of corn and cellulosic biofuel, feedstock is a large component of the total cost of production and a major source of uncertainty. According to a USDA survey, the gross feedstock costs were 80.3 cents per gallon of ethanol produce with slight differences

between small and large ethanol plants [134]. In addition to feedstock, other costs such as labor and maintenance might occur. Here, we assume that feedstock costs, labor costs, maintenance costs all grow linearly with production rate.

Let O be the total of all the constituents that scale linearly with the production rate

$$O = O_b \frac{Q}{Q_b} \tag{97}$$

where O_b is the sum of costs of all the constituents for production rate Q_b.

For the worst case scenario $T1$, the total cost can be derived from equation 96, 92 and 97.

$$Z = C + G^{T1} + O = C_b(\frac{Q}{\alpha Q_b})^s + \frac{G_t T\eta\tau(1+r)x}{2wq^{\frac{1}{2}}\alpha^{\frac{3}{2}}}Q^{\frac{3}{2}} - \frac{G_t T\eta\tau(1+r)x}{w\alpha}Q + O_b\frac{Q}{Q_b} \tag{98}$$

By minimizing the unit cost of production, an optimal production rate can be derived.

$$\min \frac{Z}{Q} = C_b(\frac{1}{\alpha Q_b})^s Q^{s-1} + \frac{G_t T\eta\tau(1+r)x}{2wq^{\frac{1}{2}}\alpha^{\frac{3}{2}}}Q^{\frac{1}{2}} - \frac{G_t T\eta\tau(1+r)x}{w\alpha} + O_b\frac{1}{Q_b} \tag{99}$$

The Karush-Kuhn-Tucker condition for an interior optimum of the above minimization problem yields the optimal size of the production system.

$$Q^*_{T1} = (\frac{G_t T\eta\tau(1+r)xQ_b^s}{4(1-s)C_b w q_c^{\frac{1}{2}}\alpha^{\frac{3}{2}-s}})^{\frac{2}{2s-3}} \tag{100}$$

Similarly for scenario $T2$, equation 96, 95 and 97 yield

$$Z = C + G^{T2} + O = C_b(\frac{Q}{\alpha Q_b})^s + \frac{G_t T\eta\tau x(2+r)}{2wq^{\frac{1}{2}}\alpha^{\frac{3}{2}}}Q^{\frac{3}{2}} - \frac{G_t T\eta\tau xr}{w\alpha}Q + O_b\frac{Q}{Q_b} \tag{101}$$

We want to minimize

$$\min \frac{Z}{Q} = C_b(\frac{1}{\alpha Q_b})^s Q^{s-1} + \frac{G_t T\eta\tau x(2+r)}{2wq^{\frac{1}{2}}\alpha^{\frac{3}{2}}}Q^{\frac{1}{2}} - \frac{G_t T\eta\tau xr}{w\alpha} + O_b\frac{1}{Q_b} \tag{102}$$

$$Q^*_{T2} = (\frac{G_t T \eta \tau x (2+r) Q_b^s}{4(1-s) C_b w q^{\frac{1}{2}} \alpha^{\frac{3}{2}-s}})^{\frac{2}{2s-3}} \tag{103}$$

B.3 Pipeline transportation cost

Liquid and gas are commonly transported through pipelines. In a processing plant, cost of piping can be 80 percent of the delivered purchased-equipment cost or 20 percent of the fixed capital investment [114]. Pipeline cost includes the capital cost of the pipe materials, labor costs, the capital and operational cost of the pump stations or compression, maintenance cost, etc. [75] [113]. The capital cost of pipeline network does not only depend on its length but also the diameters of the pipes. The operational power required for piping the chemical mixture over a given distance decreases as the diameter of the pipe increases. Developing the design for a piping system requires many considerations, such as pipe stress, ambient effects and safety. Pipeline costs can also be highly dependent on the location [113]. It is believed that piping cost scales superlinearly with pipeline length. Here, we assume that cost of piping follows the equation 104 and derive transportation cost as a function of the total production rate in the bioreactor-based fuel system described earlier.

$$G^P = G_p(\frac{D}{D_0})^z \tag{104}$$

where G_p is the unit length piping cost when the pipe length is D_0 and z is an exponent that is assumed to be 1.1 in this analysis.

Pipeline can be used to transport the chemical mixture to the central processing location continuously. Figure 25 shows an example of the pipeline layout. In this case, the total length of the pipeline is simply

$$D^{P1} = 2mX + Y = 4m^2 x + 2my = 4xm^2 + 2rxm = \frac{x}{q\alpha} Q + \frac{rx}{q^{\frac{1}{2}} \alpha^{\frac{1}{2}}} Q^{\frac{1}{2}} \tag{105}$$

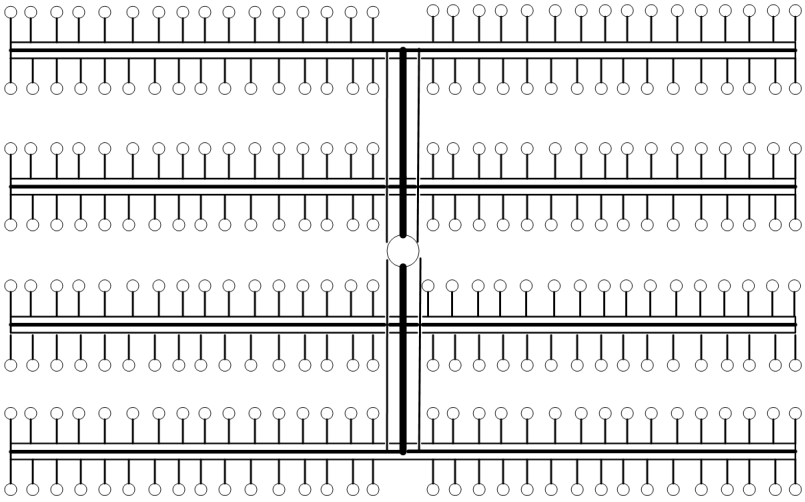


Figure 25: Baseline 3-level pipeline design

Let G_p be the unit cost of piping per distance of chemical mixture transported, the total cost of piping in this case is,

$$G^{P1} = G_p(D^{P1})^z = G_p(\frac{x}{q\alpha}Q + \frac{rx}{q^{\frac{1}{2}}\alpha^{\frac{1}{2}}}Q^{\frac{1}{2}})^z \tag{106}$$

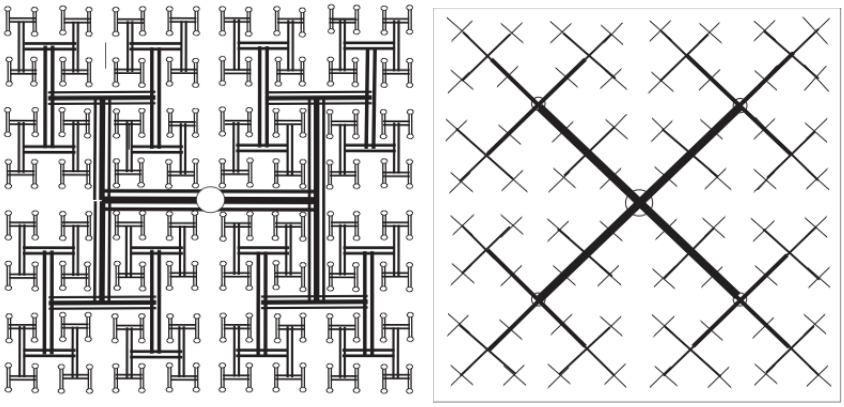


Figure 26: Scheme of a dichotomic arborescence with eight scales and a tetratomic tree, both connecting the centre point (inlet port) to 256 uniformly spaced outlet ports. Source: D. Tondeur, 2009

The pipeline system can also take other shapes, such as dichotomic arborescence or tetratomic tree, shown in Figures 26. To build a dichotomic pipeline network 26 on the same rectangular area $X \times Y$ and $m = \frac{X}{2x}$, the number of bioreactors has to satisfy $N =$

$4m^2 = 2^K$. Let l_k and x_k be the length and number of pipes at scale k

$x_k = 2^k, l_1 = X/4 = mx/2, l_2 = Y/4 = my/2, l_3 = l_1/2, l_4 = l_2/2, l_5 = l_3/2, l_6 = l_4/2$

Since $4m^2$ is even, the number of scales $K = 2\log_2 m + 2$. The total length of the pipeline network is

$$\begin{aligned} D^{P2} &= \sum_{k=1}^{K} x_k l_k = \sum_{t=1}^{\frac{K}{2}} 2^t l_1 + \sum_{t=1}^{\frac{K}{2}} 2^{t+1} l_2 \\ &= (2^{\frac{K}{2}+1} - 2)l_1 + (2^{\frac{K}{2}+2} - 2)l_2 = (2^{\frac{K}{2}} - 1)mx + (2^{\frac{K}{2}+1} - 1)my \\ &= (2m-1)mx + (4m-1)my = (2+4r)xm^2 - (1+r)xm \\ &= \frac{(1+2r)x}{2q\alpha}Q - \frac{(1+r)}{2q^{\frac{1}{2}}\alpha^{\frac{1}{2}}}Q^{\frac{1}{2}} \end{aligned} \tag{107}$$

$$G^{P2} = T_p(D^{P2})^z = G_p(\frac{(1+2r)x}{2q\alpha}Q - \frac{(1+r)}{2q^{\frac{1}{2}}\alpha^{\frac{1}{2}}}Q^{\frac{1}{2}})^z \tag{108}$$

For the tetratomic tree, $x_k = 4^k$ $l_{k+1} = \frac{1}{2}l_k$ and $l_1 = \frac{1}{4}\sqrt{X^2+Y^2} = \frac{mx\sqrt{1+b^2}}{2}$. $K = \log_2 m + 1$

$$\begin{aligned} D^{P3} &= \sum_{k=1}^{K} x_k l_k = \sum_{k=1}^{K} 4^k \frac{l_1}{2^{k-1}} = \sum_{k=1}^{K} 2^{k+1} l_1 = 2(2^{k+1}-2)l_1 \\ &= (2^{k+1}-2)mx\sqrt{1+r^2} = x\sqrt{1+r^2}(4m^2 - 2m) \\ &= \frac{x\sqrt{1+r^2}}{q\alpha}Q - \frac{x\sqrt{1+r^2}}{q^{\frac{1}{2}}\alpha}Q^{\frac{1}{2}} \end{aligned} \tag{109}$$

$$G^{P3} = G_p(D^{P3})^z = G_p(\frac{x\sqrt{1+r^2}}{q\alpha}Q - \frac{x\sqrt{1+r^2}}{q^{\frac{1}{2}}\alpha}Q^{\frac{1}{2}})^z \tag{110}$$

Instead of deriving the analytical expression, we show transportation cost as functions of production rate in the above three scenarios in the numerical analysis.

B.4 Numerical Example

In this section, we use as an example an bioreactor-based ethanol production system from algae described in [98] to compare the different transportation modes studied in this paper. Table 15 shows some basic parameter values.

Table 15: Baseline parameter values for numerical analysis

Parameter	Value
initial concentration α	1%
water density at 25 C	1000 kg/m^3
ethanol density at 25 C	789 kg/m^3
fluid density	998 kg/m^3
c_b benchmark separation cost	1.43 M \$
Q_b	1 MGY
s	0.6
Flow rate at each bioreactor q	10 m^3/yr
bioreactor length	15.24 m
bioreactor width	1.2 m

B.4.1 Trucking cost

The average operating cost of commercial trucks is estimated to be 0.65 \$ per kilometer [13] [95], including labor and maintenance costs and adjusted for cost, based on pavement roughness, driving conditions and fuel price changes. Using an average truck size of 10 tons, Figure 27 shows the result of transportation cost by trucking as a function of total flow rate in the system is given by equation 92 and 95 for scenario $T1$ and $T2$ respectively.

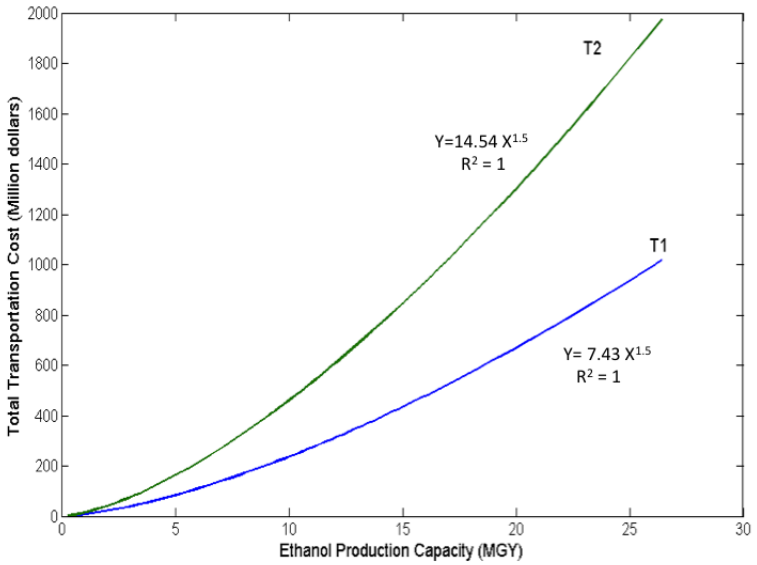


Figure 27: Total truck transport cost over 20 year project lifetime as a function of total ethanol production rate in scenario $T1$ and $T2$

As shown in figure 27, the transportation cost of scenario $T1$ is lower than that of scenario $T2$. In fact, for a fixed truck size, the number of truck tours are the same for both $T1$ and $T2$ in the same time period. However, in scenario $T2$, all trucks have to travel to the bioreactor that is at the end of the X coordinate, whereas in scenario $T1$, only the truck assigned to this bioreactor at the end of the X coordinate has to do so.

B.4.2 Pipeline cost

Here, we use flexible PVC (Polyvinyl Chloride) pipes schedule 40 as an example [149]. The price of PVC is about 1.7 $/kg as of 2010 [119]. Based on standard dimension and weight chart of PVC pipes according to ASTM D1785, we estimate a capital cost of 50 $/m for a unit length pipe. For the reference case, we plot pipeline capital cost as a function of the production rate using equation 106,108, and 110,.Figure 28 shows that the scaling exponent of capital cost of piping is 1.1 for all three scenarios considered in this analysis.

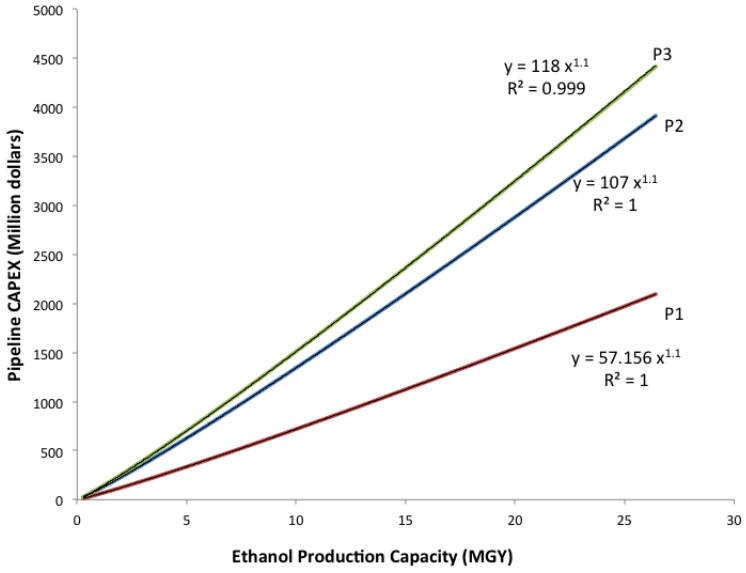


Figure 28: Pipeline capital cost as a function of total ethanol production rate in scenario $P1$, $P2$ and $P3$

REFERENCES

[1] ACADEMIES, N. R. C. O. T. N., "National research council review of the desalination and water purification technology roadmap," 2004.

[2] Administration, U. E. I., "Ranking of u.s. refineries," September 2010.

[3] Akgul, O., Zamboni, A., Bezzo, F., Shah, N., and Papageorgiou, L., "Optimization-based approaches for bioethanol supply chains," *Industrial & Engineering Chemistry Research*, 2010.

[4] Alexander, B., Barton, G., Petrie, J., and Romagnoli, J., "Process synthesis and optimisation tools for environmental design: methodology and structure," *Computers & Chemical Engineering*, vol. 24, no. 2-7, pp. 1195–1200, 2000.

[5] Anily, S. and Federgruen, A., "A class of euclidean routing problems with general route cost functions," *Mathematics of operations research*, pp. 268–285, 1990.

[6] Archetti, C., Mansini, R., and Speranza, M., "Complexity and reducibility of the skip delivery problem," *Transportation Science*, vol. 39, no. 2, pp. 182–187, 2005.

[7] Artzner, P., Delbaen, F., Eber, J., and Heath, D., "Thinking coherently," *Risk*, vol. 10, no. 11, pp. 68–71, 1997.

[8] Artzner, P., Delbaen, F., Eber, J., and Heath, D., "Coherent measures of risk," *Mathematical finance*, vol. 9, no. 3, pp. 203–228, 1999.

[9] Association, R. F., "Weekly ethanol summary," May 2011.

[10] Association, R. F., "Fuel ethanol facilities capacity by state and by plant," June 2010.

[11] Azapagic, A., "Life cycle assessment and its application to process selection, design and optimisation," *Chemical Engineering Journal*, vol. 73, no. 1, pp. 1–21, 1999.

[12] Baier, K., Duhring, U., Enke, H., Grundel, M., Lockau, W., Smith, C. R., Woods, P., Ziegler, K., Oesterhelt, C., Coleman, J., and Robert; Kramer, D., "Genetically modified photoautotrophic ethanol producing host cells, methos for producing the host cells, constructs for the transformation of the host cells, method for testing a photoautotrophic strain for desired growth property and method of producing ethanol using the host cells," 2009. PCT/EP 2009/000892.

[13] Barnes, G. and Langworthy, P., "Per mile costs of operating automobiles and trucks," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1864, no. -1, pp. 71–77, 2004.

[14] Baumgartner, K. and Thonemann, U., "Supply chain design considering economies of scale and transport frequencies," *Working paper*, 2009.

[15] Beasley, J., "Route first–cluster second methods for vehicle routing," *Omega*, vol. 11, no. 4, pp. 403–408, 1983.

[16] Bechen, K. L., "Addressing Obstacles in the Biomass Feedstock Supply Chain," 2011.

[17] Bektas, T., "The multiple traveling salesman problem: an overview of formulations and solution procedures," *Omega*, vol. 34, no. 3, pp. 209–219, 2006.

[18] Bertsimas, D. and Simchi-Levi, D., "A new generation of vehicle routing research: robust algorithms, addressing uncertainty," *Operations Research*, pp. 286–304, 1996.

[19] Bettocchi, R., Cadorin, M., Cenci, G., Morini, M., Pinelli, M., Spina, P., and Venturini, M., "Energy and Economic Analyses of Integrated Biogas-Fed Energy Systems," *Journal of Engineering for Gas Turbines and Power*, vol. 131, p. 061401, 2009.

[20] Borgonovo, E. and Peccati, L., "Uncertainty and global sensitivity analysis in the evaluation of investment projects," *International Journal of Production Economics*, vol. 104, no. 1, pp. 62–73, 2006.

[21] Boustead, I., "Eco-profiles of the european plastics industry: High density polyethylene (hdpe)," 2005.

[22] Bowling, I., Ponce-Ortega, J., and El-Halwagi, M., "Facility location and supply chain optimization for a biorefinery," *Industrial & Engineering Chemistry Research*, 2011.

[23] Brandao, J. and Mercer, A., "A tabu search algorithm for the multi-trip vehicle routing and scheduling problem," *European Journal of Operational Research*, vol. 100, no. 1, pp. 180–191, 1997.

[24] Brennan, L. and Owende, P., "Biofuels from microalgae–a review of technologies for production, processing, and extractions of biofuels and co-products," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 2, pp. 557 – 577, 2010.

[25] Broek, J., Schutz, P., Stougie, L., and Tomasgard, A., "Location of slaughterhouses under economies of scale," *European journal of operational research*, vol. 175, no. 2, pp. 740–750, 2006.

[26] Burgess, A. and Brennan, D., "Application of life cycle assessment to chemical processes," *Chemical Engineering Science*, vol. 56, no. 8, pp. 2589–2604, 2001.

[27] Cameron, J. B., Kumar, A., and Flynn, P. C., "The impact of feedstock cost on technology selection and optimum size," *Biomass and Bioenergy*, vol. 31, pp. 137–144, Feb. 2007.

[28] Carlo N Hamelinck, G. v. H. and Faaij, A. P., "Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term," *Biomass and Bioenergy*, 2005.

[29] Carpentieri, M., Corti, A., and Lombardi, L., "Life cycle assessment (LCA) of an integrated biomass gasification combined cycle (IBGCC) with CO2 removal," *Energy Conversion and Management*, vol. 46, no. 11-12, pp. 1790–1808, 2005.

[30] CCS, P. P., "Costs of pbr," August 2010.

[31] Cheng, X., DasGupta, B., and Lu, B., "Polynomial Time Approximation Scheme for Symmetric Rectilinear Steiner Arborescence Problem," *Journal of Global Optimization*, vol. 21, no. 4, pp. 385–396, 2001.

[32] Cho, J., Park, J., and Jeon, J., "Comparison of Three-and Two-Column Configurations in Ethanol Dehydration Using Azeotropic Distillation," *JOURNAL OF INDUSTRIAL AND ENGINEERING CHEMISTRY-SEOUL-*, vol. 12, no. 2, p. 206, 2006.

[33] Christofides, N., "Worst-case analysis of a new heuristic for the travelling salesman problem.," tech. rep., DTIC Document, 1976.

[34] Christofides, N., Mingozzi, A., and Toth, P., "Exact algorithms for the vehicle routing problem, based on spanning tree and shortest path relaxations," *Mathematical programming*, vol. 20, no. 1, pp. 255–282, 1981.

[35] Clarens, A., Resurreccion, E., White, M., and Colosi, L., "Environmental life cycle comparison of algae to other bioenergy feedstocks," *Environ. Sci. Technol*, vol. 44, no. 5, pp. 1813–1819, 2010.

[36] Cohen, M., "An integrated plant loading model with economies of scale and scope.," *European Journal of Operational Research*, vol. 50, no. 3, pp. 266–279, 1991.

[37] Cong, J., Kahng, A., and Leung, K., "Efficient algorithms for the minimum shortest path Steiner arborescence problem with applications to VLSI physical design," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 17, no. 1, pp. 24–39, 2002.

[38] Cong, J., Leung, K., and Zhou, D., "Performance-driven interconnect design based on distributed RC delay model," in *Design Automation, 1993. 30th Conference on*, pp. 606–611, IEEE, 2006.

[39] Corsano, G., Vecchietti, A., and Montagna, J., "Optimal design for sustainable bioethanol supply chain considering detailed plant performance model," *Computers & Chemical Engineering*, 2011.

[40] Davis, S. and Hu, P., "Transportation energy data book," *NASA STI/Recon Technical Report N*, vol. 91, p. 24681, 1991.

[41] D.O.E., U., "Natural gas 1998: Issues and trends," 1998. accessed Feb 1,2010.

[42] DOE, U., "Agricultural chemicals: Fertilizers," *Energy and Environmental Profile of the U.S. Chemical Industry*, 2000.

[43] Drezner, Z. and Hamacher, H., *Facility location: applications and theory*. Springer Verlag, 2004.

[44] Du, D., Lu, B., Ngo, H., and Pardalos, P., "Steiner tree problems," *Encyclopedia of optimization*, vol. 5, pp. 227–290, 2001.

[45] ECN, "Phyllis: The composition of biomass and waste," 2010. accessed Feb 1, 2010.

[46] Eggleston, H., "2006 ipcc guidelines for national greenhouse gas inventories," *Forestry*, vol. 5, no. OVERVIEW, pp. 1–12, 2006.

[47] Eglese, R., "Routeing winter gritting vehicles," *Discrete Applied Mathematics*, vol. 48, no. 3, pp. 231–244, 1994.

[48] Eksioglu, B., Vural, A., and Reisman, A., "The vehicle routing problem: A taxonomic review," *Computers & Industrial Engineering*, vol. 57, no. 4, pp. 1472–1483, 2009.

[49] E.P.A., U., "Catalog of chp technologies. combined heat and power partnership." accessed Feb 1, 2010.

[50] EPA, U., "Egrid database: egrid2007 version 1.1," July 2010.

[51] Feldman, E., Lehrer, F., and Ray, T., "Warehouse location under continuous economies of scale," *Management Science*, pp. 670–684, 1966.

[52] Fellows, M. and Fernau, H., "Facility location problems: a parameterized view," *Algorithmic Aspects in Information and Management*, pp. 188–199, 2008.

[53] Flynn, K., Greenwell, H., Lovitt, R., and Shields, R., "Selection for fitness at the individual or population levels: Modelling effects of genetic modifications in microalgae on productivity and environmental safety," *Journal of theoretical biology*, vol. 263, no. 3, pp. 269–280, 2010.

[54] French, R., *Open-channel hydraulics*. McGraw Hill Book Co., New York, NY, 1985.

[55] Fu, L., "Scheduling dial-a-ride paratransit under time-varying, stochastic congestion," *Transportation Research Part B: Methodological*, vol. 36, no. 6, pp. 485–506, 2002.

[56] Gallagher, P., Brubaker, H., and Shapouri, H., "Plant size: Capital cost relationships in the dry mill ethanol industry," *BIOMASS & BIOENERGY*, vol. 28, no. 6, pp. 565–571, 2005.

[57] Gan, J. and Smith, C., "Optimal plant size and feedstock supply radius: A modeling approach to minimize bioenergy production costs," *Biomass and Bioenergy*, 2010.

[58] Gan, J. and Smith, C., "Coupling greenhouse gas credits with biofuel production cost in determining conversion plant size," *SILVA FENNICA*, vol. 44, no. 3, pp. 497–510, 2010.

[59] Garey, M. and Johnson, D., *Computers and Intractability: A Guide to the Theory of NP-completeness*. WH Freeman & Co., 1979.

[60] Geiver, L., "Houston algae developer to release electromechanical processor," 2011.

[61] Ghafoori, E., Flynn, P., and Feddes, J., "Pipeline vs. truck transport of beef cattle manure," *Biomass and Bioenergy*, vol. 31, no. 2-3, pp. 168–175, 2007.

[62] Goldman, J., "Outdoor algal mass cultures–II. Photosynthetic yield limitations," *Water research*, vol. 13, no. 2, pp. 119–136, 1979.

[63] Greenwell, H., Laurens, L., Shields, R., Lovitt, R., and Flynn, K., "Placing microalgae on the biofuels priority list: a review of the technological challenges," *Journal of The Royal Society Interface*, vol. 7, no. 46, p. 703, 2010.

[64] Guillard, R., "Culture of phytoplankton for feeding marine invertebrates," *Culture of marine invertebrate animals*, pp. 26–60, 1975.

[65] Guinée, J., "Handbook on life cycle assessment operational guide to the iso standards," *The international journal of life cycle assessment*, vol. 7, no. 5, pp. 311–313, 2002.

[66] Guirardello, R. and Swaney, R., "Optimization of process plant layout with pipe routing," *Computers & chemical engineering*, vol. 30, no. 1, pp. 99–114, 2005.

[67] Guirardello, R. and Swaney, R. E., "Optimization of process plant layout with pipe routing," *Computers & Chemical Engineering*, vol. 30, pp. 99–114, Nov. 2005.

[68] Haimovich, M. and Kan, A., "Bounds and heuristics for capacitated routing problems," *Mathematics of Operations Research*, pp. 527–542, 1985.

[69] Hajiaghayi, M., Mahdian, M., and Mirrokni, V., "The facility location problem with general cost functions," *Networks*, vol. 42, no. 1, pp. 42–47, 2003.

[70] Homma, T. and Saltelli, A., "Importance measures in global sensitivity analysis of nonlinear models," *Reliability Engineering & System Safety*, vol. 52, no. 1, pp. 1–17, 1996.

[71] Hrastel, I., Gerbec, M., and Stergaršek, A., "Technology optimization of wet flue gas desulfurization process," *Chemical Engineering & Technology*, vol. 30, no. 2, pp. 220–233, 2007.

[72] Huang, Y., Bird, R., and Heidrich, O., "Development of a life cycle assessment tool for construction and maintenance of asphalt pavements," *Journal of Cleaner Production*, vol. 17, no. 2, pp. 283–296, 2009.

[73] Hwang, F., "On Steiner minimal trees with rectilinear distance," *SIAM Journal on Applied Mathematics*, vol. 30, no. 1, pp. 104–114, 1976.

[74] I.E.A., U., "Solar paces task i: Solar thermal electric systems." accessed Feb 1, 2010.

[75] Institute, T. H., "Pumping lifecycle costs: A guide to lcc analysis for pumping systems," tech. rep., U.S. Department of Energy, January 2001.

[76] Jack, M. W., "Scaling laws and technology development strategies for biorefineries and bioenergy plants," *Bioresource Technology*, vol. 100, pp. 6324–6330, Dec. 2009.

[77] Jaramillo, P., Griffin, W., and Matthews, H., "Comparative life-cycle air emissions of coal, domestic natural gas, LNG, and SNG for electricity generation," *Environ. Sci. Technol*, vol. 41, no. 17, pp. 6290–6296, 2007.

[78] Jenkins, B., "A comment on the optimal sizing of a biomass utilization facility under constant and variable cost scaling," *Biomass and Bioenergy*, vol. 13, no. 1-2, pp. 1–9, 1997.

[79] Johnson, D., "A note on dijkstra's shortest path algorithm," *Journal of the ACM (JACM)*, vol. 20, no. 3, pp. 385–388, 1973.

[80] Junginger, M., De Visser, E., Hjort-Gregersen, K., Koornneef, J., Raven, R., Faaij, A., and Turkenburg, W., "Technological learning in bioenergy systems," *Energy Policy*, vol. 34, no. 18, pp. 4024–4041, 2006.

[81] Kang, S., Onal, H., Ouyang, Y., Scheffran, J., and Tursun, U., "Optimizing the biofuels infrastructure: Transportation networks and biorefinery locations in illinois," *Handbook of Bioenergy Economics and Policy*, pp. 151–173, 2010.

[82] Kelleher, T. and James, R., "Distillation studies in a high-gravity contactor," *Industrial & engineering chemistry research*, vol. 35, no. 12, pp. 4646–4655, 1996.

[83] Khumawala, B. and Kelly, D., *Warehouse location with concave costs*. Herman C. Krannert Graduate School of Industrial Administration, Purdue University (Lafayette, Ind), 1972.

[84] Klose, A. and Drexl, A., "Facility location models for distribution system design," *European Journal of Operational Research*, vol. 162, no. 1, pp. 4–29, 2005.

[85] Kongshaug, G., "Energy consumption and greenhouse gas emissions in fertilizer production," in *IFA technical conference, Marrakech, Morocco*, vol. 28, 1998.

[86] Kumar, A., Cameron, J., and Flynn, P., "Pipeline transport of biomass," *Applied Biochemistry and Biotechnology, 113*, vol. 1, no. 3, pp. 27–39, 2004.

[87] Kumar, A., Cameron, J., and Flynn, P., "Biomass power cost and optimum plant size in western Canada," *Biomass and Bioenergy*, vol. 24, no. 6, pp. 445–464, 2003.

[88] Kunjapur, A. and Eldridge, R., "Photobioreactor Design for Commercial Biofuel Production from Microalgae," *Industrial & Engineering Chemistry Research*, vol. 49, no. 8, pp. 3516–3526, 2010.

[89] Kuppan, T., *Heat exchanger design handbook*. CRC, 2000.

[90] Kwiatkowski, J., McAloon, A., Taylor, F., and Johnston, D., "Modeling the process and costs of fuel ethanol production by the corn dry-grind process," *Industrial crops and products*, vol. 23, no. 3, pp. 288–296, 2006.

[91] Laporte, G., "The vehicle routing problem: An overview of exact and approximate algorithms," *European Journal of Operational Research*, vol. 59, no. 3, pp. 345–358, 1992.

[92] Laporte, G. and Nobert, Y., "A cutting planes algorithm for the m-salesmen problem," *Journal of the Operational Research Society*, pp. 1017–1023, 1980.

[93] Lardon, L., Hélias, A., Sialve, B., Steyer, J., and Bernard, O., "Life-cycle assessment of biodiesel production from microalgae," *Environmental science & technology*, vol. 43, no. 17, pp. 6475–6481, 2009.

[94] Lardon, L., Hlias, A., Sialve, B., Steyer, J.-P., and Bernard, O., "Life-cycle assessment of biodiesel production from microalgae," *Environmental Science and Technology*, vol. 43, pp. 6475–6481, May 2009.

[95] Levinson, D., Corbett, M., and Hashami, M., "Operating costs for trucks," *Working Papers*, 2004.

[96] Linde, D., *CRC Handbook of Chemistry and Physics, 90th ed*. CRC:Boca Raton: CRC Press, 2009.

[97] Lombardi, L., "Life cycle assessment comparison of technical solutions for CO2 emissions reduction in power generation," *Energy conversion and management*, vol. 44, no. 1, pp. 93–108, 2003.

[98] Luo, D., Hu, Z., Choi, D., Thomas, V., Realff, M., and Chance, R., "Life Cycle Energy and Greenhouse Gas Emissions for an Ethanol Production Process Based on Blue-Green Algae," *Environmental science & technology*, pp. 117–146, 2010.

[99] Luo, D., Hu, Z., Choi, D. G., Thomas, V., Realff, M., and Chance, R., "Lifecycle energy and greenhouse gas emissions for an ethanol production process based on blue-green algae," *Environmental Science and Technology*, October 2010.

[100] Luterbacher, J., Fröling, M., Vogel, F., Maréchal, F., and Tester, J., "Hydrothermal Gasification of Waste Biomass: Process Design and Life Cycle Asessment," *Environmental science & technology*, vol. 43, no. 5, pp. 1578–1583, 2009.

[101] Magalhaes, A. and others, "Techno-economic assessment of biomass preconversion processes as a part of biomass-to-liquids line-up," *Biofuels, Bioproducts and Biorefining*, vol. 3, no. 6, pp. 584–600, 2009.

[102] McAloon, "Determining the cost of producing ethanol from corn starch and lignocellulosic feedstocks," 2000.

[103] McCool, B., "Personal communication with Ben McCool, Algenol Inc.," 2011.

[104] Mele, F., Guillén-Gosálbez, G., and Jiménez, L., "Optimal planning of supply chains for bioethanol and sugar production with economic and environmental concerns," *Computer Aided Chemical Engineering*, vol. 26, pp. 997–1002, 2009.

[105] Melis, A., "Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency," *Plant science*, vol. 177, no. 4, pp. 272–280, 2009.

[106] Melo, M., Nickel, S., and Saldanha-Da-Gama, F., "Facility location and supply chain management-a review," *European Journal of Operational Research*, vol. 196, no. 2, pp. 401–412, 2009.

[107] Mirchandani, P. and Francis, R., *Discrete location theory*, vol. 23. Wiley-Interscience, 1990.

[108] Nastansky, L., Selkow, S., and Stewart, N., "Cost-minimal trees in directed acyclic graphs," *Mathematical Methods of Operations Research*, vol. 18, no. 1, pp. 59–67, 1974.

[109] Nguyen, M. and Prince, R., "A simple rule for bioenergy conversion plant size optimisation: bioethanol from sugar cane and sweet sorghum," *Biomass and Bioenergy*, vol. 10, no. 5-6, pp. 361–365, 1996.

[110] Overend, R., "The average haul distance and transportation work factor for biomass delivered to a central plant," *Biomass*, vol. 2, pp. 75–79, 1982.

[111] Owen, S. and Daskin, M., "Strategic facility location: A review," *European Journal of Operational Research*, vol. 111, no. 3, pp. 423–447, 1998.

[112] Ozyurt, Z. and Aksen, D., "Solving the multi-depot location-routing problem with lagrangian relaxation," *Extending the horizons: Advances in computing, optimization, and decision technologies*, pp. 125–144, 2007.

[113] Parker, N., "Using natural gas transmission pipeline costs to estimate hydrogen pipeline costs," *Institute of Transportation Studies, UC Davis*, 2004.

[114] Peters, M. S. and Timmerhaus, K. D., *Chemical Engineering Series: Plant Design and Economics for Chemical Engineers*. McGraw-Hill, Inc, fifth edition ed., 2003.

[115] Pflug, G., "Some remarks on the value-at-risk and the conditional value-at-risk," *Probabilistic constrained optimization: Methodology and applications*, vol. 38, pp. 272–281, 2000.

[116] Phimister, J. and Seider, W., "Semicontinuous, middle-vessel, extractive distillation," *Computers & Chemical Engineering*, vol. 24, no. 2-7, pp. 879–885, 2000.

[117] Plus, A., "Aspen Plus User Guide," *Aspen Technology Limited, Cambridge, Massachusetts, United States*, 2003.

[118] Posten, C. and Schaub, G., "Microalgae and terrestrial biomass as source for fuels–A process view," *Journal of biotechnology*, vol. 142, no. 1, pp. 64–69, 2009.

[119] Pricing, I., "Icis pricing: Polyvinyl chloride (usa)," May 2010.

[120] Pricing, I., "Ethanol prices and pricing information," May 2011.

[121] Rego, C. and Roucairol, C., "Using tabu search for solving a dynamic multi-terminal truck dispatching problem," *European Journal of Operational Research*, vol. 83, no. 2, pp. 411–429, 1995.

[122] Release, E. P., "Synthetic Genomics Inc. and Exx- onMobil Research and Engineering Company Sign Exclusive, Multi-Year Agreement to Develop Next Generation Biofuels Using Photosynthetic Algae," 2009.

[123] ReVelle, C. and Eiselt, H., "Location analysis: A synthesis and survey," *European Journal of Operational Research*, vol. 165, no. 1, pp. 1–19, 2005.

[124] Revelle, C., Eiselt, H., and Daskin, M., "A bibliography for some fundamental problem categories in discrete location science," *European Journal of Operational Research*, vol. 184, no. 3, pp. 817–848, 2008.

[125] Rockafellar, R. and Uryasev, S., "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, pp. 21–42, 2000.

[126] Romeijn, H., Shu, J., and Teo, C., "Designing two-echelon supply networks," *European Journal of Operational Research*, vol. 178, no. 2, pp. 449–462, 2007.

[127] Sahin, G. and Sural, H., "A review of hierarchical facility location models," *Computers & Operations Research*, vol. 34, no. 8, pp. 2310–2331, 2007.

[128] Saltelli, A., "Making best use of model evaluations to compute sensitivity indices," *Computer Physics Communications*, vol. 145, no. 2, pp. 280–297, 2002.

[129] Saltelli, J. and others, "Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques," *Reliability Engineering & System Safety*, vol. 28, no. 2, pp. 229–253, 1990.

[130] Sawitzki, D., "Implicit flow maximization by iterative squaring," *SOFSEM 2004: Theory and Practice of Computer Science*, pp. 83–94, 2003.

[131] Searcy, E. and Flynn, P., "The impact of biomass availability and processing cost on optimum size and processing technology selection," *Applied biochemistry and biotechnology*, vol. 154, no. 1, pp. 92–107, 2009.

[132] Sears, S., "Vapor compression distillation system and method," 1999. US Patent 5,968,321.

[133] Sears, S., "Vapor Compression Distillation: Adding High Tech Understanding to Nature's Process," *WATER CONDITIONING AND PURIFICATION INTERNATIONAL*, p. 26, 2006.

[134] Shapouri, H., Gallagher, P., of Agriculture. Office of Energy Policy, U. S. D., and Uses, N., *USDA's 2002 ethanol cost-of-production survey*. United States Department of Agriculture, Office of the Chief Economist, Office of Energy Policy and New Uses, 2005.

[135] Sheehan, J., Aden, A., Paustian, K., Killian, K., Brenner, J., Walsh, M., and Nelson, R., "Energy and environmental aspects of using corn stover for fuel ethanol," *Journal of Industrial Ecology*, vol. 7, no. 3-4, pp. 117–146, 2003.

[136] Sheehan, J. and others, *A look back at the US Department of Energy's Aquatic Species Program: Biodiesel from algae*, vol. 328. National Renewable Energy Laboratory, 1998.

[137] Shi, W. and Su, C., "The rectilinear Steiner arborescence problem is NP-complete," in *Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms*, pp. 780–787, Society for Industrial and Applied Mathematics, 2000.

[138] Sibley, L., "Dow, algenol to build pilot algae-based biorefinery," 2009. accessed Feb 1,2010.

[139] Singh, J. and Gu, S., "Commercialization potential of microalgae for biofuels production," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 9, pp. 2596 – 2610, 2010.

[140] Singh, R. P. and Mahar, P. S., "Optimal design of multidiameter, multioutlet pipelines," *Journal of Water Resources Planning and Management*, vol. 129, pp. 226–233, May 2003.

[141] Skone, T. J. and Gerdes, K. J., "Petroleum-based fuels life cycle greenhouse gas analysis 2005 baseline model," 2009.

[142] Skone, T. and Gerdes, K., "Petroleum-based fuels life cycle greenhouse gas analysis - 2005 baseline model," 2009.

[143] Sobol, I., "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Mathematics and Computers in Simulation*, vol. 55, no. 1-3, pp. 271–280, 2001.

[144] Soland, R., "Optimal facility location with concave costs," *Operations Research*, pp. 373–382, 1974.

[145] Solomon, B., Barnes, J., and Halvorsen, K., "Grain and cellulosic ethanol: History, economics, and energy policy," *BIOMASS & BIOENERGY*, vol. 31, pp. 416–425, June 2007.

[146] Thompson, C., *A complexity theory for VLSI*. PhD thesis, Citeseer, 1980.

[147] Tondeur, D. and Kvaalen, E., "Equipartition of entropy production. An optimality criterion for transfer and separation processes," *Industrial & Engineering Chemistry Research*, vol. 26, no. 1, pp. 50–56, 1987.

[148] Tondeur, D. and Luo, L., "Design and scaling laws of ramified fluid distributors by the constructal approach," *Chemical Engineering Science*, vol. 59, pp. 1799–1813, apr 2004.

[149] Toolbox, T. E., "Standard dimensions and weight of pvc pipes according astm d1785," July 2011.

[150] Toth, P. and Vigo, D., *The vehicle routing problem*, vol. 9. Society for Industrial Mathematics, 2002.

[151] Tursun, U., Kang, S., Onal, H., Ouyang, Y., and Scheffran, J., "Optimal biorefinery locations and transportation network for the future biofuels industry in illinois," in *Environmental and Rural Development Impacts Conference, October 15-16, 2008, St. Louis, Missouri*, Farm Foundation, Transition to a Bio Economy Conferences, 2008.

[152] USDA, "Weekly ethanol summary," May 2011.

[153] U.S.E.P.A., "Renewablefuelstandardprogram.finalrule," 2010. accessed Feb 1, 2010.

[154] Uslu, A., Faaij, A., and Bergman, P., "Pre-treatment technologies, and their effect on international bioenergy supply chain logistics. techno-economic evaluation of torrefaction, fast pyrolysis and pelletisation," *Energy*, vol. 33, no. 8, pp. 1206–1223, 2008.

[155] Vane, L. and Alvarez, F., "Membrane-assisted vapor stripping: energy efficient hybrid distillation–vapor permeation process for alcohol–water separation," *Journal of Chemical Technology & Biotechnology*, vol. 83, no. 9, pp. 1275–1287, 2008.

[156] Vane, L., Alvarez, F., Huang, Y., and Baker, R., "Experimental validation of hybrid distillation-vapor permeation process for energy efficient ethanol–water separation," *Journal of Chemical Technology & Biotechnology*, vol. 85, no. 4, pp. 502–511, 2010.

[157] Vazirani, V., *Approximation algorithms*. Springer Verlag, 2001.

[158] Wakeley, H., Hendrickson, C., Griffin, W., and Matthews, H., "Economic and environmental transportation effects of large-scale ethanol production and distribution in the united states," *Environmental science & technology*, vol. 43, no. 7, pp. 2228–2233, 2009.

[159] Wang, M., "Greenhouse gases, regulated emissions, and energy use in transportation (greet) model, version 1.8 b," *Argonne National Laboratory*, 2007.

[160] Wechsatol, W., Lorente, S., and Bejan, A., "Tree-shaped flow structures with local junction losses," *International Journal of Heat and Mass Transfer*, vol. 49, no. 17-18, pp. 2957–2964, 2006.

[161] Wechsatol, W., Lorente, S., and Bejan, A., "Tree-shaped flow structures with local junction losses," *International Journal of Heat and Mass Transfer*, vol. 49, pp. 2957–2964, Aug. 2006.

[162] West, G. and Brown, J., "The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization," *JOURNAL OF EXPERIMENTAL BIOLOGY*, vol. 208, pp. 1575–1592, May 2005.

[163] Weyer, K., Bush, D., Darzins, A., and Willson, B., "Theoretical maximum algal oil production," *BioEnergy Research*, vol. 3, no. 2, pp. 204–213, 2010.

[164] Williams, P. and Laurens, L., "Microalgae as biodiesel & biomass feedstocks: Review & analysis of the biochemistry, energetics & economics," *Energy & Environmental Science*, vol. 3, no. 5, pp. 554–590, 2010.

[165] Woelfel, P., "Symbolic topological sorting with OBDDs," *Journal of Discrete Algorithms*, vol. 4, no. 1, pp. 51–71, 2006.

[166] Woods, P., "Algenol biofuel's direct to ethanol technology," 2009.

[167] Woods, P., Legere, E., Moll, B., Unamunzaga, C., and Mantecon, E., "Closed photobioreactor system for continued daily in situ production, separation, collection, and removal of ethanol from genetically enhanced photosynthetic organisms," Mar. 23 2010. US Patent 7,682,821.

[168] Wright, C., Pryfogle, P., Stevens, N., Hess, J., and Radtke, C., "Value of distributed preprocessing of biomass feedstocks to a bioenergy industry," *system*, vol. 9, p. 12, 2006.

[169] Wright, M. and Brown, R., "Establishing the optimal sizes of different kinds of biorefineries," *BIOFUELS BIOPRODUCTS & BIOREFINING-BIOFPR*, vol. 1, pp. 191–200, Nov. 2007.

[170] Wu, L., Zhang, X., and Zhang, J., "Capacitated facility location problem with general setup cost," *Computers & operations research*, vol. 33, no. 5, pp. 1226–1241, 2006.

[171] Wu, M., Wu, Y., and Wang, M., "Energy and emission benefits of alternative transportation liquid fuels derived from switchgrass: a fuel life cycle assessment," *Biotechnology Progress*, vol. 22, no. 4, pp. 1012–1024, 2006.

[172] Xie, W. and Sahinidis, N., "A branch-and-bound algorithm for the continuous facility layout problem," *Computers & Chemical Engineering*, vol. 32, no. 4-5, pp. 1016–1028, 2008.

[173] Xu, G. and Papageorgiou, L., "A construction-based approach to process plant layout using mixed-integer optimization," *Ind. Eng. Chem. Res*, vol. 46, no. 1, pp. 351–358, 2007.

[174] Xu, G. and Papageorgiou, L., "Process plant layout using an improvement-type algorithm," *Chemical Engineering Research and Design*, vol. 87, no. 6, pp. 780–788, 2009.

[175] Yang, C. and Ogden, J., "Determining the lowest-cost hydrogen delivery mode," *INTERNATIONAL JOURNAL OF HYDROGEN ENERGY*, vol. 32, pp. 268–286, Feb. 2007.

[176] You, F. and Wang, B., "Life cycle optimization of biomass-to-liquids supply chains with distributed-centralized processing networks," *Industrial & Engineering Chemistry Research*, 2011.

[177] Youssef, M. and Mahmoud, M., "An iterative procedure for solving the uncapacitated production-distribution problem under concave cost function," *International Journal of Operations & Production Management*, vol. 16, no. 3, pp. 18–27, 1996.

[178] Zamboni, A., Shah, N., and Bezzo, F., "Spatially explicit static model for the strategic design of future bioethanol production systems. 1. cost minimization," *Energy & Fuels*, vol. 23, no. 10, pp. 5121–5133, 2009.

VITA

Dexin Luo was born on January 3, 1984 in Jining, Shandong Province, China. Her mother Yaling Sun is a pharmacist and her father Shuying Luo was a mechanical engineer. After attending Jining No. 1 High School, Dexin attended Qingdao University in China, earning a Bachelor's degree in Environmental Science in 2005. In the following year, she studied and worked as a research assistant in the Environmental Systems Institute in Tsinghua University, China. In January of 2007, Dexin joined the School of Industrial and Systems Engineering at Georgia Institute of Technology to pursue a Ph.D. in Industrial Engineering under the supervision of Professor Valerie M. Thomas and Professor Matthew J. Realff. She subsequently received her Master of Science degree in Industrial Engineering at Georgia Tech in 2008. Dexin's hobbies include sports (swimming and basketball), music and traveling.