A prototype simulation-based optimization approach to model feedstock development for chemical process industry

Ismail Fahmi, Selen Cremaschi*

University of Tulsa, Oklahoma, 800 S Tucker Drive, Tulsa, OK 74104, USA

Abstract

Incorporating non-traditional feedstocks, e.g., biomass, to chemical process industry (CPI) will require investments in research & development (R&D) and capacity expansions. The impact of these investments on the evolution of biomass to commodity chemicals (BTCC) system should be studied to ensure a cost-effective transition with acceptable risk levels. The BTCC system includes both exogenous, e.g., product demands (decision-independent) and endogenous, e.g., the change in technology cost with investment levels (decision-dependent) uncertainties. This paper presents a prototype simulation-based optimization (SIMOPT) approach to study the BTCC system evolution under exogenous and endogenous uncertainties, and provides a preliminary analysis of the impact of using three different sampling methods, i.e., Monte Carlo, Latin Hypercube, and Halton sequence, to generate the simulation runs on the computational cost of the SIMOPT approach. The results of a simplified case study suggest that annual demand increases is the dominant factor for the total cost of the BTCC system. The results also suggest that using Halton sequence as the sampling method yields the smallest number of samples, i.e., the least computational cost, to achieve a statistically significant solution.

Keywords: Simulation-based optimization, Biomass to commodity chemicals; Exogenous and endogenous uncertainty; Latin Hypercube sampling; Halton sequence

1. Introduction

CPI, a large consumer of fossil fuels, will have to find alternative feedstocks to support the market demand in the future (Dodds and Gross, 2007). Biomass is a promising alternative because it is abundant, locally available, and renewable. To realize the shift of using biomass as the feedstock to supplement and/or instead of fossil-based feedstocks for the CPI will require substantial capital and R&D investments to bring the biomass technologies to commercial availability.

The decisions of how much to invest in which technologies for the short, medium and long term in the resource-constrained environment of the CPI is a challenging problem due to its strong combinatorial character and several sources of uncertainty. The main sources of uncertainty are: (1) many possible evolution paths of the technologies from their current maturation levels to their deployment, (2) the yields of the technologies after deployment at the time of the initial capital investments, (3) the demand and raw material composition and availability. The uncertainties in the evolution paths and the yields of the technologies are endogenous uncertainty, i.e., the decisions may impact their realizations and/or distribution parameters, and the demand and raw material related uncertainties are exogenous uncertainties, i.e., they are decision-independent (Goel and Grossmann, 2006). Therefore, the resulting problem is a stochastic optimization (SP) problem under endogenous and exogenous uncertainties.

The literature on stochastic optimization problems with endogenous uncertainty is limited and can be grouped based on whether the resolution of uncertainty (Colvin and Maravelias, 2008, 2009; Goel and Grossmann, 2004, 2005; Goel et al., 2006; Held and Woodruff, 2005; Jonsbråten et al., 1998; Mercier and Van Hentenryck, 2008; Tarhan and Grossmann, 2008; Tarhan et al., 2009) or the distribution parameters of uncertainty (Ahmed, 2000; Held and Woodruff, 2005; Mercier and Van Hentenryck, 2008; Viswanath et al., 2004) are impacted by the decisions.

In the first group of studies, the resolution of the uncertainty is decision-dependent. Jonsbråten et al. (Jonsbråten et al., 1998) is one of the first to study this problem for MILPs. The authors developed a branch-and-bound algorithm coupled with an implicit enumeration method to generate the scenario trees. Investment and operational planning of gas field developments, in which the revelation of uncertainty in gas reserves depends on the investment decisions, was studied by Goel and Grossmann (Goel and Grossmann, 2004). The authors formulated the problem as a multi-stage stochastic MILP, and proposed an approximation algorithm based on decomposition and search space restrictions to solve it. A Lagrangean duality based branch-and-bound algorithm, which is guaranteed to give the optimal solution, was proposed to solve a similar formulation (Goel and Grossmann, 2005; Goel et al., 2006). Tarhan et al. (Tarhan et al., 2009) expanded the MILP of Goel and Grossmann (Goel and Grossmann, 2004) to incorporate nonlinear reservoir models and gradual uncertainty revelation. The resulting large-scale mixed integer nonlinear programming (MINLP) was solved using a duality-based branch-and-bound algorithm. Mercier and Van Hentenryck (Mercier and Van Hentenryck, 2008) developed "Anytime Multi-Step Anticipatory Algorithm (Amsaa) to solve stochastic optimization problems with both exogenous and endogenous uncertainty. The algorithm uses the sampling average approximation method to approximate the problem as a Markov Decision Process (MDP), heuristic search algorithm to solve the MDP, and stochastic combinatorial optimization solvers to compute good upper bounds for the heuristic search algorithm. Tarhan and Grossmann, (Tarhan and Grossmann, 2008) investigated the impact of the value of building pilot plants prior to major plant investments for the synthesis of process networks. They modeled the problem as MILP with gradual uncertainty resolution with time and proposed a duality-based branch-and-bound algorithm for solving the problem. Colvin and Maravelias (Colvin and Maravelias, 2008, 2009, 2010) incorporated the decision dependent revelation of uncertainty in clinical trial outcomes explicitly in their SP formulations.

The literature that deals with the second class of problems, where the decisions impact the distribution parameters of the uncertainty, is scarce. Ahmed (Ahmed, 2000) formulated single stage stochastic optimization problems with discrete decision and endogenous uncertainty as 0-1 hyperbolic programs. The effect of decisions on the uncertainty distribution parameters were modeled using utility function. They generated MILP equivalents of the 0-1 hyperbolic programs, and solved the resulting MILP using linear programming (LP) based branch-and-bound algorithms. Viswanath et al. (Viswanath et al., 2004) developed a two-stage stochastic program to minimize the expected shortest distance between start and end nodes of a network under random failures. The survival probability of a link in the network could be increased by investing on it at the first stage. They approximated the objective using the first order terms of the multilinear function obtained by relaxing the integrality constraints of the investment decision variables and using Taylor series expansion, and used an iterative solution algorithm to improve the approximate objective function value. Held and Woodruff (Held and Woodruff, 2005) introduced the problem of multi-stage interdiction of networks, where the structure of the network is uncertain. The objective was to maximize the probability of interdicting the flow of information or goods through the network, and probability distribution of the next stage network structure depended on the interdiction decisions. The authors investigated the use of heuristic approaches.

Rigorous SP approaches become computationally intractable with the size of the problem and source of uncertainties because the scenario and decision trees grow exponentially. To date, the largest stochastic optimization problem with endogenous uncertainties solved, which was an R&D pipeline management problem, included seven products (Colvin and Maravelias, 2010), and it only considered the uncertainty in the outcome of the clinical trials. Considering the number of technologies that should be evaluated for the BTCC investment planning problem combined with the evolution path and demand uncertainties, the problem cannot be solved to optimality with the rigorous SP approaches with the current state-of-art algorithms. However, the simulation-based optimization (SIMOPT) approach with its ability to accommodate both exogenous and endogenous uncertainty in the simulation and its ability to react to endogenous uncertainty with recourse actions with the re-optimization provides an excellent framework for the problem under consideration.

The SIMOPT approach uses the concept of timelines to generate multiple, unique realizations of the controlled evolution of the system under study (Subramanian et al., 2003). A typical SIMOPT timeline involves several applications of the Deterministic Optimization (DO) algorithm to determine optimized values for system degrees of freedom. In between each DO on a timeline, the simulation is used to determine the integrated system behavior that accounts for uncertainty. The various timelines arise from randomness introduced through changes in parameter values, and exceptional random events that are imposed on the simulation. In the outer loop, information gathered from different timelines is used to perform risk analysis or to update the parameters of deterministic optimization modules used within the inner loop (Subramanian et al., 2003). There are two main sources of computational cost for SIMOPT approach: (1) the number of samples, i.e., timelines, required to have a statistically significant coverage of the uncertain parameters space and (2) the computational cost of repetitive DO problem solution.

In this paper, we present a prototype SIMOPT framework to study the impact of technology evolution paths and demand uncertainties on the BTCC system evolution, and use this SIMOPT framework to investigate the effect of different sample generation methods on the uncertain parameter space coverage and on the overall SIMOPT computational cost. The sampling methods considered are pure random sampling (Monte Carlo Simulations), Latin Hypercube sampling, and sampling based on Halton sequences. The paper is organized as follows: Following section gives the problem description. Next, the details of the prototype SIMOPT approach developed for this work is presented including the details of the DO and the BTCC simulation. A brief overview of the considered sampling approaches follows. The case study, its results, and the results of the sampling method comparisons are provided in Section 4. Finally, the conclusions and future directions are presented.

2. Problem statement

Given biomass and fossil-based feedstock processing technologies and their characteristics, the initial commodity chemicals production system market conditions, and system uncertainties, the objective is to study the BTCC system evolution under endogenous and exogenous uncertainties. Two sources of uncertainties are considered in this study. One is learning elasticity uncertainties, the parameter used to express how fast a technology cost decreases with the amount of total capacity and R&D expenditure increases; this uncertainty is the endogenous one. The second one is the demand for the products, and it is the exogenous uncertainty.

3. A prototype simulation-based optimization framework for BTCC evolution

The developed SIMOPT framework (Figure 1) to study the possible evolution paths of the BTCC system (i.e., different timelines) combines the deterministic optimization (DO) and stochastic simulation models of the BTCC system. The DO model determines the capacity expansion ($X_{e,t}$), R&D expenditure ($RD_{e,t}$), and production ($P_{e,t}$) decisions given the current state of the system, the expected demand profile, and expected learning elasticities. The cumulative capacity ($CX_{e,t}$), total R&D expenditure ($CRD_{e,t}$), unit expansion cost ($CC_{e,t}$), unit raw material cost ($CR_{v,t}$), demand ($D_{v,t}$), and total cost ($TotCostSim_t$) at the current simulation time, t, defines the current state of the system. At the beginning of each timeline (i.e., when t = 0), the deterministic optimization solution for the initial market conditions and the mean of the uncertain parameters is used. The stochastic simulation models the behavior of the BTCC system with the suggested investment decisions, and under demand and learning elasticity uncertainties. The simulation returns the control to the optimization module when the realization of the elasticity factor and demand uncertainties result in infeasibility in the original investment policy suggested by the optimization, i.e., in case of trigger events. Two trigger events modeled in this work are: At any time t, **(1)** the total cumulative capacity is not enough to support the demand, and **(2)** the total realized cost (the one calculated by the simulation, $TotCostSim_t$) is higher than the predicted total cost (the one predicted by DO, $TotCostOpt_t$) with an assumed tolerance level. The optimization is then used to determine the optimal investment policy using the current state of the BTCC system, the estimated learning elasticities, and demand forecasts for the remainder of the planning horizon. The iteration between the simulation and optimization continues until the end of the planning horizon. The information generated for one planning horizon or timeline represents an example of how the BTCC system could evolve. Many unique evolutions of BTCC system result due to different realizations of uncertainty and resulting investment decisions. We study the generated timelines to understand the impact of uncertainties on the total BTCC-system cost variation and technology selections.

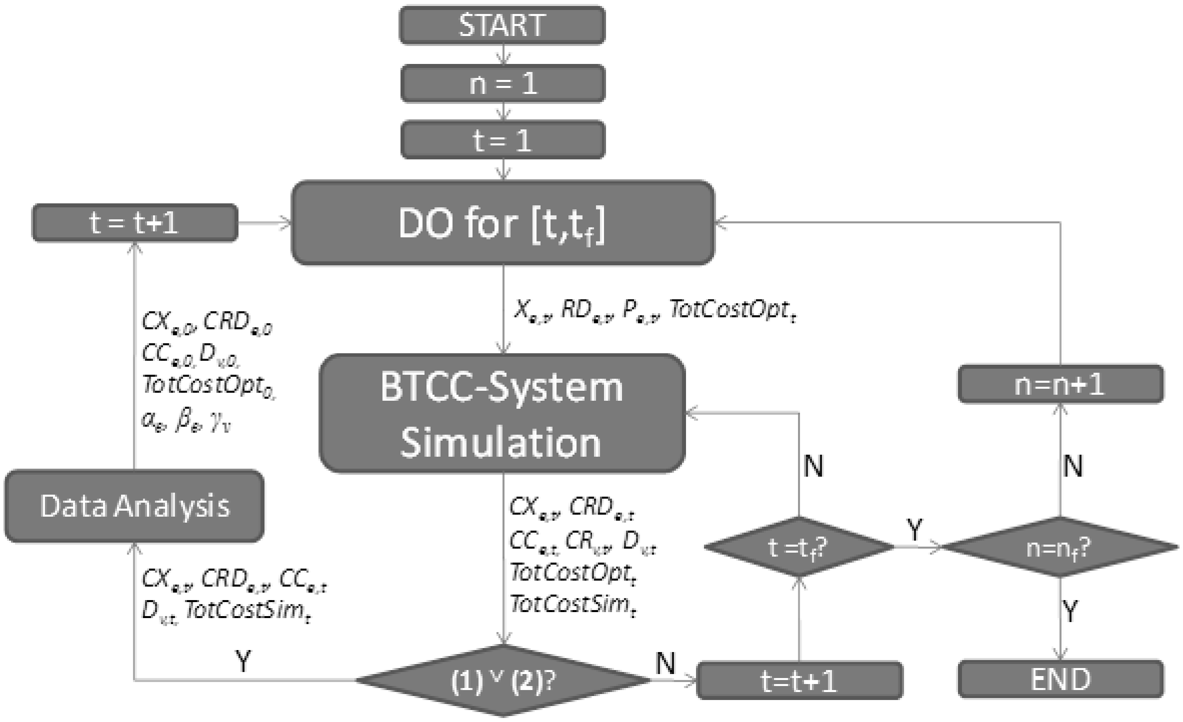


Figure 1: The developed SIMOPT approach

3.1. Deterministic optimization module

3.1.1. Formulation Tools for BTCC Investment Planning Problem

The Network Representation, the Two-factor Learning Curve and the Stage-gate Representation are used to represent the BTCC system. This section briefly describes the details of each formulation tool.

3.1.1.1. The Network Representation

The main purpose of the Network Representation is to keep track of the material flow through all interconnecting technologies (Cremaschi, 2011). In the Network Representation, the nodes, v, and the directed-arcs, e, represent the materials and the processing technologies, respectively (Figure 2) (Cremaschi, 2011). The overall interconnection of the technologies in the BTCC system is expressed with a weighted incidence matrix, B, that utilizes the yield of each technology, η_e (Cremaschi, 2011). Matrix B has the dimensions of $|V|\times|E|$ and the definition of each element of this matrix is given in Eqn (1).

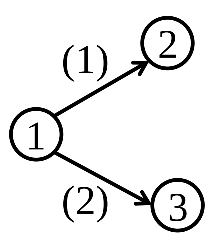


Figure 2: A simple network representation with three materials (nodes) and two technologies (directed-arcs)

$$b_{v,e} = \begin{cases} -1/\eta_e & \text{if material } v \text{ is a raw material for technology } e \\ 1 & \text{if material } v \text{ is produced by technology } e \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

3.1.1.2. Two-factor Learning Curve

The Two-factor Learning Curve (Kouvaritakis et al., 2004) expresses the relationship between the technology expansion cost, $CC_{e,t}$, the total installed production capacity, $CX_{e,t}$, and the total R&D expenditure, $CRD_{e,t}$ for each technology e at each time t. The mathematical expression for this relationship is given in Eqn (2). α_e and β_e are referred to as the learning-by-doing and learning-by-searching elasticity, respectively. Please note that these two values are negative as the technology cost decreases when the cumulative capacity or the R&D expenditure increases.

$$CC_{e,t} = CC_{e,0} \left(\frac{CX_{e,t}}{CX_{e,0}} \right)^{\alpha_e} \left(\frac{CRD_{e,t}}{CRD_{e,0}} \right)^{\beta_e} \qquad \forall t, e \tag{2}$$

3.1.1.3. The Stage-gate Representation

The Stage-gate Representation uses an analogous concept to the pharmaceutical R&D pipelines described in Blau et al. (Blau et al., 2004) and the Technology Readiness Level Metric by Sadin et al. (Sadin et al., 1989). The purpose of the Stage-gate Representation is to allow the explicit expression of the technology maturity stage (Fahmi and Cremaschi, 2011). Realistically, a technology will have to have a certain installed capacity (i.e., a certain maturity level) to support the overall market needs (Fahmi and Cremaschi, 2011).

We define a four level metric (consistent with the traditional chemical engineering technology development process.): (1) research stage, (2) pilot plant stage, (3) advancement stage, and (4) commercial stage. A representing schema is given in Figure 3.

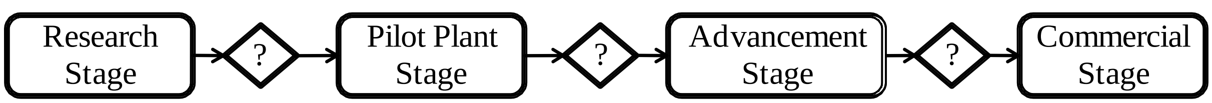


Figure 3: Stage-gate framework for technology development

The Stage-gate Representation utilizes a binary variable $Y_{e,s,t}$ to define the maturity level, i.e., the evolution stage, of a technology (Eqn (3)). The subscript s represents the defined four-level metric, (i.e. stages (1) - (4)):

$$Y_{e,s,t} = \begin{cases} 1 & \text{if technology } e \text{ is at least at stage } s \text{ at time } t \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

3.1.2. MINLP formulation

The tools explained in Section 3.1 are then utilized to formulate the MINLP for the BTCC system. The objective and the constraints of the optimization problem are detailed in Figure 4.

Objective Function

$$\text{Min}\, TC$$

Subject to

Cost Function

$$TC = \sum_e \sum_t CC_{e,t}\,(CX_{e,t} - CX_{e,t-1}) + \sum_{v \in VR} \sum_t CR_{v,t}\, R_{v,t} + \sum_e \sum_t RD_{e,t} \qquad (4)$$

Technology Costs

$$CC_{e,t} = CC_{e,0} \left(\frac{CX_{e,t}}{CX_{e,0}}\right)^{\alpha_e} \left(\frac{CRD_{e,t}}{CRD_{e,0}}\right)^{\beta_e} \quad \forall t, e \qquad (5)$$

Raw Material Costs

$$CR_{v,t} = CR_{v,0} + k_v \sum_{j=1}^{t} R_{v,j}\,(1 + IR)^t \quad \forall t, \{v \mid v \in VR \wedge v \notin VRR\} \qquad (6)$$

$$CR_{v,t} = CR_{v,0}\,(1 + IR)^t \quad \forall t, v \in VRR \qquad (7)$$

Product Demands

$$D_{v,t} = D_{v,0}\,(1 + \gamma_v)^t \quad \forall t, v \in VP \qquad (8)$$

Meet Product Demands

$$R_{v,t} \geq D_{v,t} \quad \forall t, v \in VP \qquad (9)$$

$$R_{v,t} = \sum_e b_{v,e}\, P_{e,t} \quad \forall t, v \in VP \qquad (10)$$

No Accumulation of Intermediates

$$\sum_e b_{v,e}\, P_{e,t} = 0 \quad \forall t, \{v \mid v \notin VP \wedge v \notin VR\} \qquad (11)$$

Raw Material Requirements

$$R_{v,t} = \sum_e - b_{v,e}\, P_{e,t} \quad \forall t, v \in VR \qquad (12)$$

Capacity Constraints

$$\begin{bmatrix} Y_{e,3,t} \\ P_{e,t} \leq CX_{e,t} \end{bmatrix} \vee \begin{bmatrix} \neg Y_{e,3,t} \\ P_{e,t} = 0 \end{bmatrix} \quad \forall t, e \qquad (13)$$

Capacity and *R&D* Stock Bounds

$$CX_{e,t-1} \leq CX_{e,t} \quad \forall t, e \qquad (14)$$

$$CRD_{e,t-1} \leq CRD_{e,t} \quad \forall t, e \qquad (15)$$

Interstage Preevolution Requirements

$$Y_{e,s,t} \leq Y_{e,s-1,t} \quad \forall t, e, \{s \mid s \in \{2,3,4\}\} \qquad (16)$$

No Back Evolution

$$Y_{e,s,t} \geq Y_{e,s,t-1} \quad \forall t, e, s \qquad (17)$$

Stage Bracketing

$$\begin{bmatrix} Y_{e,s,t} \\ CX_{e,t} \geq LO_s \end{bmatrix} \vee \begin{bmatrix} \neg Y_{e,s,t} \\ CX_{e,t} \geq 0 \end{bmatrix} \quad \forall t, e, s \qquad (18)$$

$$CX_{e,t} \leq \max(HI_s) \quad \forall t, e, \{s \mid Y_{e,s,t}\} \qquad (19)$$

Nomenclature

TC : Total cost
$CC_{e,t}$: Unit capital cost for technology e at time t
$CX_{e,t}$: Cumulative installed capacity of technology e at time t
$P_{e,t}$: Amount of production with technology e at time t
$CR_{v,t}$: Unit cost of material v at time t
$R_{v,t}$: Amount of material v produced or consumed at time t
$CRD_{e,t}$: Total R&D expenditure for technology e at time t
$Y_{e,s,t}$: A binary variable, defined as 1 if technology e is at least at stage s at time t and 0 otherwise
k_v : Constant cost increase coefficient for material v (defined only for nonrenewable raw materials)
IR : Inflation rate
$D_{v,t}$: Demand for material v at time t (defined only for products)
γ_v : Annual increasing rate of demand for material v
LO_s : The lower limit of the cumulative capacity for stage s
HI_s : The upper limit of the cumulative capacity for stage s
VR : Raw material set
VRR : Renewable raw material set (subset of VR)
VP : Products set

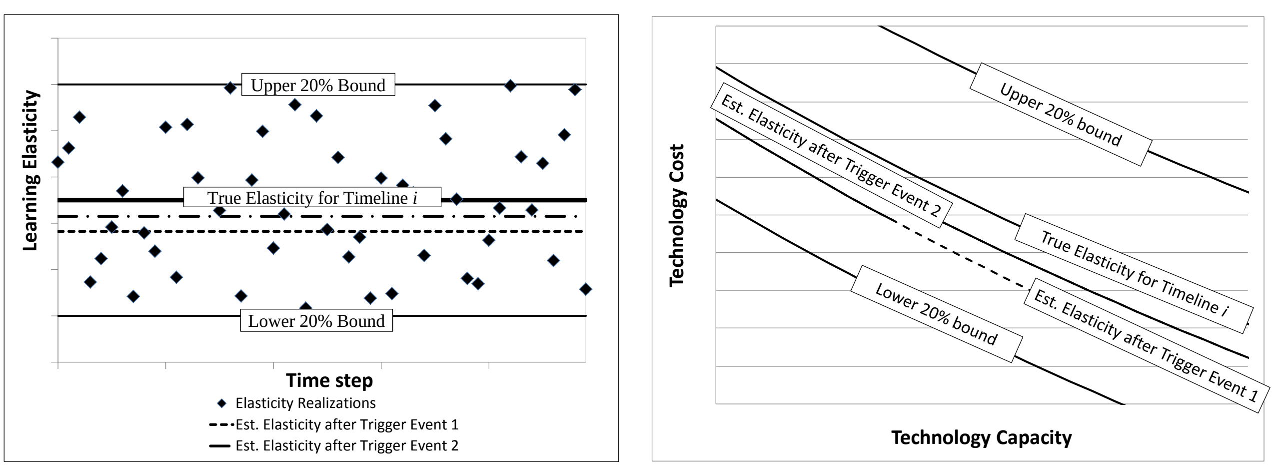
Figure 4: The MINLP formulation of BTCC technologies evolution

The objective is to minimize Total Cost (TC). Eqn (4) calculates this value by summing over the raw material cost, the technology expansion cost, and the R&D expenditure. Eqn (5) expresses the technology cost relationship according to the two-factor learning curve model. Eqns (6) and (7) express the relationship of raw material costs for non-renewable and renewable feedstock, respectively. In this model, it is assumed that product demands grow with an annual increasing rate of γ_v and this is represented in Eqn (8). Eqns (9) and (10) ensures that the production satisfies the market demand. Eqn (11) states that intermediate materials do not accumulate. Eqn (12) computes the amount of required raw material. Eqn (13) uses the state-gate representation to express that only technologies with the third and above maturity level can contribute to the production. Eqns (14) and (15) state that cumulative capacity and total R&D expenditure cannot decrease with time. Eqn (16) makes sure that technologies do not skip stages. Eqn (17) prohibits back evolution (i.e., a technology reverting back to a lower maturity stage). Finally, Eqn (18) expresses the relationship between cumulative capacity and maturity stage.

3.2. Simulation module

The inputs for the BTCC-system simulation (modeled in ExtendSim AT V8) are the capacity expansions, R&D expenditures, and the production schedule of each technology (e), and predicted total cost at each time step (t) (Figure 1). For each time step, the simulation calculates the realized cumulative capacity for each technology, total R&D expenditure, unit expansion cost of each technology, production, unit raw material cost, and total cost under demand and learning elasticity uncertainties.

The product demand for each time step is computed according to Eqn (8), where γ_v (the exogenous uncertainty) follows a normal distribution. For each technology, the learning curve elasticities, α_e and β_e (endogenous uncertainties), are modeled with a two-tier probability distribution. The first tier, which is assumed to follow a beta or normal distribution, represents the different unique realization of learning elasticities for each timeline. Modeling the learning elasticity uncertainties of undeveloped technologies with beta distributions ensure that their realized values stay between -1.0 and 0.0. The second tier (upper and lower bounds of the elasticity in Figure 5a), which is assumed to follow a uniform distribution, mimics the two-factor learning curve modeling errors for each time step within a single timeline. In other words the realized elasticity values are within these bounds with an average of the true elasticity for each time step within the same timeline (Figure 5a). Figure 5b shows the relationship between the technology cost and capacity when the realized learning elasticities are used to calculate these values with a single-factor learning curve model.



(a) (b)

Figure 5: The illustration of the two-tiered probability distribution application to represent the endogenous uncertainties (a) Sample elasticity factor calculations within a timeline (b) The resulting technology vs capacity curve using the elasticity factor

At each time step, the BTCC-system simulation: (a) calculates the cumulative capacity and total R&D expenditure for each technology using their corresponding schedules, (b) estimates the available capacity for each technology based on their maturity level (a technology must at least be at the Advancement Stage to contribute to production), (c) calculates the unit expansion cost for each technology using two-factor learning curve with the realized α_e and β_e, (d) computes total available production for each product, (e) generates a trigger event if the available production capacity is less than the realized demand (if a trigger event is generated, the simulation stops at the end of that time step and calls optimization), (f) calculates the unit production cost of each production route for each product using raw material costs, (g) generates a priority list – based on unit production costs – that is utilized to decide which production route will be selected to fulfill each product demand if there is more than one production route (in ascending order of unit production costs), (h) assigns production to each technology according to the priority list and realized demand, (i) compares realized product demand with the realized production, (j) computes product surplus and the corresponding storage costs, or backorder amount, (k) calculates realized total cost as the summation of realized capacity expansion costs, R&D expenditure and production costs, and (l) generates a trigger event if the realized total cost is higher than the (1 + tolerance level) × predicted total cost or lower than the (1 - tolerance level) × predicted total cost.

3.3. Data analysis module

Data analysis module calculates the learning elasticities, α_e and β_e, and the demand forecasts. Given $CC_{e,t}$, $CX_{e,t}$, and $CRD_{e,t}$ from time zero to time t for each technology e, the learning elasticities for technology e are estimated by nonlinear-regression using the two-factor learning curve relationship. An illustration of when a new learning elasticity is obtained after one or more trigger events can be seen in Figure 5. Notice that as the system progresses through the time steps, the estimated elasticity becomes closer to the true value (Figure 5). The value of γ_v in demand equation is estimated via nonlinear-regression using realized demands for each product v ($D_{v,t}$) from time zero to time t.

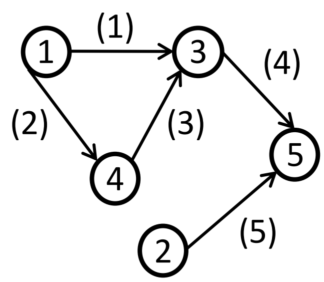
4. An overview of considered sampling methods

The objective of this study is to investigate the evolution of the BTCC system under uncertainties. To incorporate uncertainties to the system, a simulation-optimization (SIMOPT) framework is developed. One of the major computational costs for SIMOPT framework is the required number of timelines to effectively sample the uncertain parameter space. Therefore, one needs to find a sampling method that will require as few timelines as possible, but still obtain a pre-set statistical significance. In this work, three sampling methods, Monte Carlo sampling (MCS), the Latin Hypercube sampling (LHS), and the Halton sampling, are used to propagate the uncertainties in the developed simulation-based optimization framework. With MCS, each sample point is generated randomly according to their distributions for each uncertain parameter. With LHS, each uncertain parameter is segmented into equal probability intervals, and a sample point is generated randomly from each interval for each parameter. Then, the samples of uncertain parameter intervals are paired randomly. The basic idea of LHS is to cover the uncertain space more uniformly than MCS. With Halton sampling, instead of generating random values, a specific mathematical series is used to generate the samples (Lee and Chen, 2009).

To compare the performance of these sampling methods, we used bootstrapping method to compute the upper and lower 95% confidence intervals (CIs) of the mean and standard deviation of the total production cost, and to determine the number of timelines it took for these values to stabilize. It is expected that the CIs of these values will not change once the uncertain parameter space is covered adequately. With MCS and Halton sampling approaches, the upper and lower 95% CIs can be calculated at increments of one timeline because the patterns of MCS is random and Halton samples follows a sequence (Lee and Chen, 2009) where all sample points from a sample with a size of m can be reused to generate a sample with a size of $m+1$. For this study, each bootstrap batch (n+1) has an increment of 50 timelines from the previous batch (n) for MCS and Halton sampling. However, LHS does not yield itself easily to incremental sampling studies. In general, to obtain a sample size of m+1 realizations with LHS pattern from a sample size of m realizations, which also has LHS pattern, requires *m*+1 additional points. Nuchitprasittichai and Cremaschi (Nuchitprasittichai and Cremaschi, 2012) developed an incremental LHS (iLHS) algorithm that minimizes the number of additional points required while keeping the LHS pattern. This algorithm was utilized here to minimize the computational cost of LHS approach, and the increments were set to 500 samples, and therefore, each bootstrap batch for LHS method has an increment of 500 timelines, instead of 50. For each sampling approach, the maximum number of sample points was set to 5000.

5. Case study

Ethylene production case given in Cremaschi (Cremaschi, 2011) is used in this study. In this case study, ethylene can be produced from both a non-renewable feedstock, naphtha, and a renewable feedstock, lignocellulosic biomass. The naphtha is cracked to produce ethylene. The biomass can be fermented to produce ethanol, and then dehydrogenated to produce ethylene, or alternatively, gasified to syngas, then converted to ethanol, and dehydrogenated to ethylene. The network representation and the details of the nodes and arcs can be seen in Figure 6. A technology is assumed to be in the Development Stage above 10 million tons production capacity, and the tolerance level for the total cost is assumed to be 20%. The distribution parameters of the uncertain variables are given in Table 1. These uncertain parameters include the learning elasticities for all five technologies and the annual demand rate increase of the final product.



Nodes	Edges
1 → Biomass	(1) → Fermentation
2 → Naphtha	(2) → Gasification
3 → Ethanol	(3) → Catalytic conversion
4 → Syngas	(4) → Catalytic dehydration
5 → Ethylene	(5) → Cracking

Figure 6: Network representation of the ethylene production case study (Cremaschi 2011)

The values of the model parameters at t=0 are given in Table 2. The solution of the optimization problem using these parameter values is the first solution passed to simulation at t=0 for all timelines. This initial optimum solution of the DO dictates a major investment for capacity expansion of the fermentation-dehydration path and at t=43 supports investments on the gasification-conversion-dehydration path (Cremaschi, 2011). The total cost of this solution is US$ 30.92 trillion.

6. Results and discussion

6.1. Comparison of the sampling methods

Figure 7 shows the change in the upper and lower bounds of 95% CIs for the mean and the standard deviation of the total BTCC-system production cost with the number of SIMOPT timelines calculated using Monte Carlo (Figure 7 a-b), Halton (Figure 7 c-d), and Latin Hypercube (Figure 7 e-f) sampling methods. For the 95% CI of the mean obtained using the bootstrapping method, MCS reaches a value of US$182 billion at 2850 timelines, Halton sampling reaches US$170 billion at 1600 timelines, and LHS never gets lower than US$219 billion at 3500 timelines. We selected the mentioned number of timelines for MCS and Halton sampling for the comparison of the absolute difference of the 95% CI of the mean because at these numbers of timelines, the values of mean CIs are stabilized. The 95% CI of the standard deviation obtained using the bootstrapping method for MCS, Halton sampling, and LHS reach US$210 billion, US$152 billion, and US$230 billion at 5000 timelines, respectively. To put these values in perspective, the mean total cost of the system is around US$33 trillion and the mean standard deviation is around $3 trillion. These results show that for the SIMOPT approach of the BTCC system for this case study, Halton sampling method gives the lowest number of sampling points for the same uncertain space coverage, hence the lowest computational cost from the aspect of uncertain parameter sampling, compared to the other two approaches.

Table 1: Distribution parameters of the ethylene production case study

	1st Tier Dist.	**2nd Tier Dist.**
α_1	Beta(14.1, 46.2)	U(0.8x, 1.2x)
α_2	Beta(11.1, 46.2)	U(0.8x, 1.2x)
α_3	Beta(14.1, 46.2)	U(0.8x, 1.2x)
α_4	Beta(14.1, 46.2)	U(0.8x, 1.2x)
α_5	N(0, 0.003)	U(x-0.2, x+0.2)
γ_{eth}	N(0.026,0.003)	N/A
β_1	Beta(2.5, 21.7)	U(0.8x, 1.2x)
β_2	Beta(1.7, 15.3)	U(0.8x, 1.2x)
β_3	Beta(2.5, 21.7)	U(0.8x, 1.2x)
β_4	Beta(2.5, 21.7)	U(0.8x, 1.2x)
β_5	N(0, 0.003)	U(x-0.2, x+0.2)

Table 2: Deterministic parameters of the ethylene production case study

	Value
α_1	-0.21
α_2	-0.28
α_3	-0.21
α_4	-0.21
α_5	0
γ_{eth}	0.026
β_1	-0.07
β_2	-0.05
β_3	-0.07
β_4	-0.07
β_5	0

6.2. Analysis of the solution

In this section, we will discuss the results obtained using the Halton sampling method with 1600 timelines. The mean and the standard deviation of the total production cost are US$32.93 trillion and US$2.87 trillion, respectively. The total BTCC-system costs range from US$30.72 trillion to US$46.99 trillion. As can be seen

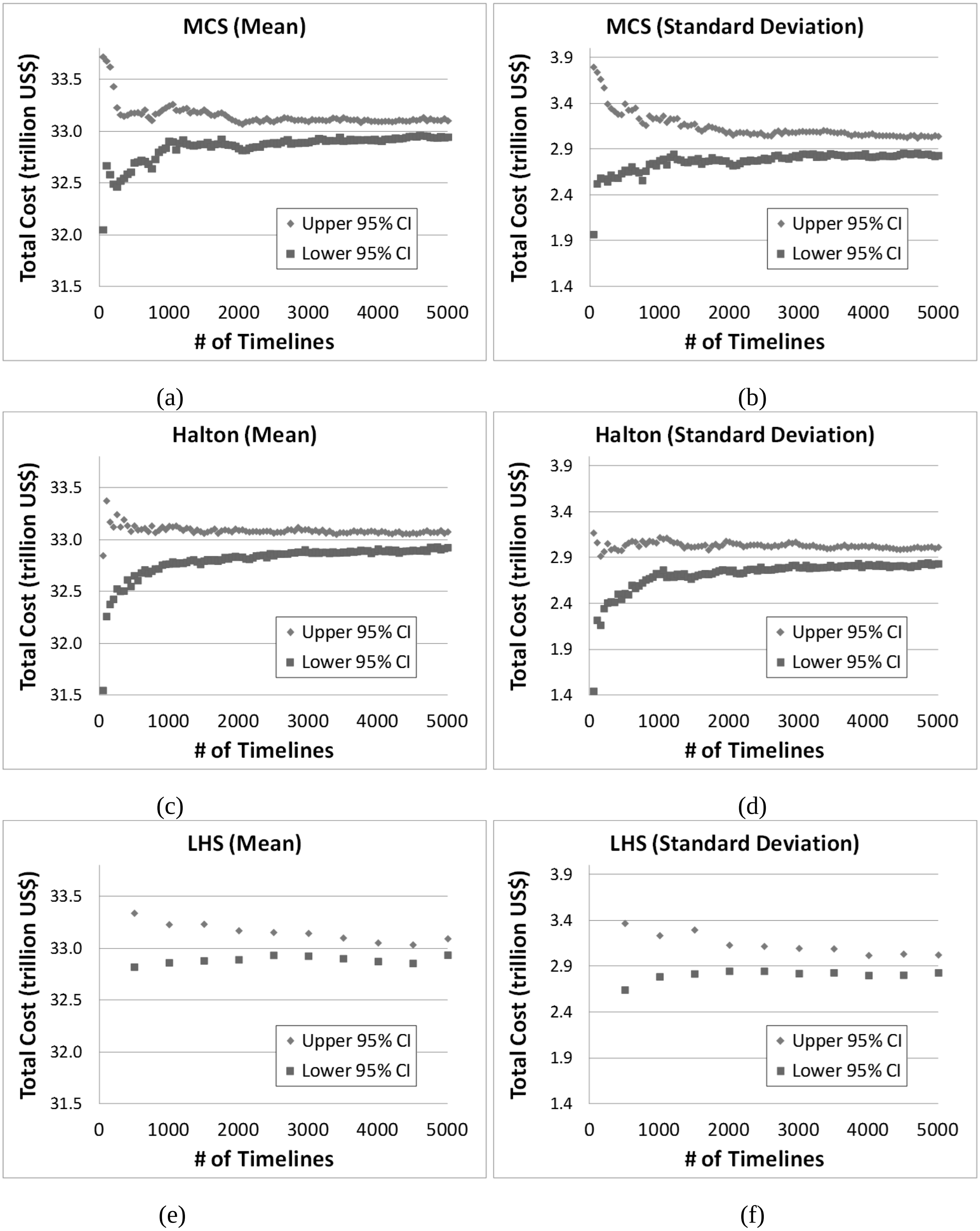


Figure 7: The mean and the standard deviation of the total cost computed using the bootstrapping method

1.68) and is very sharp (with a kurtosis of 2.69). There are two main BTCC evolution paths yielding this total cost behavior. In the first case, the system does not generate any trigger events, i.e., optimization is not recalled. Therefore, the initial DO solution at t=0 is implemented until the end of planning horizon. This solution recommends a major investment on fermentation-dehydration path and an eventual minor investment on the gasification-conversion-dehydration path for transitioning into using biomass as a feedstock. The total costs for these cases range from US$30.72 trillion to US$46.99 trillion. The change in the total cost for these cases is the result of different elasticity factor realizations. This case corresponds to 1121 timelines out of 1600, and hence explains 70.04% of the timelines. In the second case, the solution recommends to only invest in the fermentation-dehydration path to incorporate the biomass feedstock to produce ethylene. Because no investment on the gasification-conversion-dehydration was realized, it deviates slightly from the initial DO solution. The total costs for these cases range from US$30.72 trillion to US$43.82 trillion. The driving force for the deviation from the initial DO solution is the product demand forecast. When the realized demand is higher than the predicted one in DO, utilizing and further expanding the technologies with already-high capacity becomes more attractive than trying to expand the underdeveloped technologies in order to lower their costs. This case corresponds to 479 timelines, hence explains 29.96% of the timelines. Therefore, the immediate recommendation from this study would be to invest in fermentation-dehydration path, and to follow the demand carefully to decide whether to invest on a gasification-conversion-dehydration path on a future date.

A closer look at the timelines reveal that the dominant factor for the total cost is the annual rate of demand increase, i.e., the greater the annual rate of increase, the higher the production cost. An example of this can be seen in minimum and maximum total production costs. The minimum total cost corresponds to the annual rate of demand increase realization of 0.0249, while the maximum total cost corresponds to the annual rate of demand increase realization of 0.0345. This result highlights the importance of accurate demand forecasting for reducing the total cost uncertainties for transitioning to biomass as a supplement or replacement feedstock for chemical process industry.

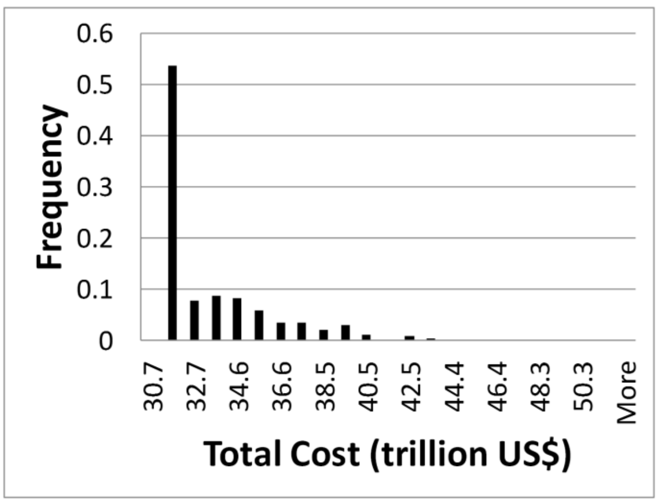


Figure 8: The total cost distribution using Halton sampling method with 1600 timelines

7. Conclusions and Future Directions

In order to study the impact of exogenous and endogenous uncertainties on the BTCC system evolution, a prototype simulation-based optimization approach is presented. This approach is successful in producing many unique realizations of the BTCC system evolution resulting from different investment schedules due to different realization of the uncertainties. Monte Carlo, Latin Hypercube, and Halton sampling methods were applied to cover the uncertain parameters space for the BTCC investment planning problem. The results show that the dominant factor for determining BTCC-system total cost is the annual rate of demand increase suggesting accurate demand forecasting as a robust way to reduce BTCC-system total cost. Based on our case study results, it was observed that Halton sampling method covers the uncertain parameter space with least number of timelines to get a statistically significant result for the BTCC system evolution.

For future directions, other sampling methods will be explored to determine which method covers the uncertain parameter space accurately with the least number of samples. Some possibilities include the univariate dimension reduction and polynomial chaos expansion.

8. Acknowledgements

Financial support from NSF CAREER Award No 1055974 and a research grant to use the ExtendSim AT simulation tool from Imagine That Inc. are greatly acknowledged.

9. References

Ahmed, S., 2000. Strategic planning under uncertainty: Stochastic integer programming approaches, PhD Dissertation in Industrial Engineering. University of Illinois at Urbana-Champaign.

Blau, G.E., Pekny, J.F., Varma, V.A., Bunch, P.R., 2004. Managing a Portfolio of Interdependent New Product Candidates in the Pharmaceutical Industry. Journal of Product Innovation Management 21, 227-245.

Colvin, M., Maravelias, C.T., 2008. A stochastic programming approach for clinical trial planning in new drug development. Computers & Chemical Engineering 32, 2626-2642.

Colvin, M., Maravelias, C.T., 2009. Scheduling of testing tasks and resource planning in new product development using stochastic programming. Computers & Chemical Engineering 33, 964-976.

Colvin, M., Maravelias, C.T., 2010. Modeling methods and a branch and cut algorithm for pharmaceutical clinical trial planning using stochastic programming. European Journal of Operational Research 203, 205-215.

Cremaschi, S., 2011. Modeling Next Generation Feedstock Development for Chemical Process Industry. Computer Aided Chemical Engineering 29, 1040-1044.

Dodds, D.R., Gross, R.A., 2007. Chemicals from Biomass. Science 318, 1250-1251.

Fahmi, I., Cremaschi, S., 2011. Stage-gate Representation of Feedstock Development for Chemical Process Industry. Foundations of Computer-Aided Process Operations 2012

Goel, V., Grossmann, I.E., 2004. A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves. Computers & Chemical Engineering 28, 1409-1429.

Goel, V., Grossmann, I.E., 2005. A lagrangean duality based branch and bound for solving linear stochastic programs with decision dependent uncertainty. Computer Aided Chemical Engineering Volume 20, 55-60.

Goel, V., Grossmann, I.E., 2006. A Class of stochastic programs with decision dependent uncertainty. Mathematical Programming 108, 355-394.

Goel, V., Grossmann, I.E., El-Bakry, A.S., Mulkay, E.L., 2006. A novel branch and bound algorithm for optimal development of gas fields under uncertainty in reserves. Computers & Chemical Engineering 30, 1076-1092.

Held, H., Woodruff, D., 2005. Heuristics for Multi-Stage Interdiction of Stochastic Networks. Journal of Heuristics 11, 483-500.

Jonsbråten, T., Wets, R., Woodruff, D., 1998. A class of stochastic programs withdecision dependent random elements. Annals of Operations Research 82, 83-106.

Kouvaritakis, N., Soria, A., Isoard, S., 2004. Modelling energy technology dynamics: methodology for adaptive expectations models with learning by doing and learning by searching. International Journal of Global Energy Issues 14, 104-115.

Lee, S., Chen, W., 2009. A comparative study of uncertainty propagation methods for black-box-type problems. Structural and Multidisciplinary Optimization 37, 239-253.

Mercier, L., Van Hentenryck, P., 2008. Amsaa: A Multistep Anticipatory Algorithm for Online Stochastic Combinatorial Optimization Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems. 5015, 173-187.

Nuchitprasittichai, A., Cremaschi, S., 2012. An algorithm to determine sample sizes for optimization with artificial neural networks (in press). AIChE Journal, DOI: 10.1002/aic.13871.

Sadin, S.R., Povinelli, F.P., Rosen, R., 1989. The NASA technology push towards future space mission systems. Acta Astronautica 20, 73-77.

Subramanian, D., Pekny, J.F., Reklaitis, G.V., Blau, G.E., 2003. Simulation-optimization framework for stochastic optimization of R&D pipeline management. AIChE Journal 49, 96-112.
Tarhan, B., Grossmann, I.E., 2008. A multistage stochastic programming approach with strategies for uncertainty reduction in the synthesis of process networks with uncertain yields. Computers & Chemical Engineering 32, 766-788.
Tarhan, B., Grossmann, I.E., Goel, V., 2009. Stochastic Programming Approach for the Planning of Offshore Oil or Gas Field Infrastructure under Decision-Dependent Uncertainty. Industrial & Engineering Chemistry Research 48, 3078-3097.
Viswanath, K., Peeta, S., Salman, S.F., 2004. Investing in the Links of a Stochastic Network to Minimize Expected Shortest Path Lengths. Institute for Research in the Behavioral, Economics, and Management Sciences, 2004. Paper No. 1167.