

A Data-Driven Optimization Approach for the Optimization of Shale Gas Supply Chains under Uncertainty

Jiyao Gao, Fengqi You*

Cornell University, 318 Olin Hall, Ithaca, New York 14853, USA
fengqi.you@cornell.edu

This paper aims to leverage the big data in shale gas industry for better decision making in optimal design and operations of shale gas supply chains under uncertainty. We propose a two-stage distributionally robust optimization model, where uncertainties associated with both the upstream shale well estimated ultimate recovery and downstream market demand are simultaneously considered. In this model, decisions are classified into first-stage design decisions as well as second-stage operational decisions. A data-driven approach is applied to construct the ambiguity set based on principal component analysis and first-order deviation functions. By taking advantage of affine decision rules, a tractable mixed-integer linear programming formulation can be obtained. The applicability of the proposed modeling framework is demonstrated through a case study of Marcellus shale gas supply chain. Comparisons with deterministic optimization models are investigated as well.

1. Introduction

Most existing supply chain design and optimization studies rely on centralized models (Garcia and You, 2015). In other words, a single decision maker is assumed to oversee the whole supply chain (Aiello et al., 2017). Thus, all the decisions can be implemented successfully to pursue a single objective (Asala et al., 2017). However, supply chains in practice are normally managed by multiple stakeholders, and each stakeholder may pursue different objectives. Thus, the conflict of interest may eventually lead to compromised solutions that are not in favour of any single stakeholder (Cachon and Netessine, 2004). Consequently, the optimal solutions obtained from centralized models can be suboptimal or even infeasible in a decentralized supply chain. To explicitly address the interest of each stakeholder in supply chain optimization problems, multiple game theoretic models are developed (Gao and You, 2017a). For instance, there are models for the optimization of cooperative multi-enterprise supply chains based on the generalized Nash bargaining solution approach (Yue and You, 2014a). On the other hand, optimization models integrating game theories of Stackelberg game and Nash-equilibrium are proposed for the design and operations of non-cooperative supply chains (Gao and You, 2017b). However, one important assumption made in these models is that the information among stakeholders is deterministic. In other words, all the stakeholders make decisions based on perfect information of the supply chain (Yue and You, 2014b). This assumption may not hold in real-world applications since there are always time delays between the decisions, and uncertainties are ubiquitous in the supply chain decision making process (Gong and You, 2017). Different uncertainty realizations may significantly affect the rational behaviours of stakeholders. Therefore, it is considered imperative to develop a holistic game theoretic model for systematic optimization of supply chains with different stakeholders considering their uncertainty behaviours.

Motivated by this knowledge gap, a novel modelling framework is proposed in this work, where the leader-follower Stackelberg game (Stackelberg, 2010) is integrated with two-stage stochastic programming approach to formulate a holistic game theoretic model. Compared with traditional game theoretic models, this modelling framework allows consideration of uncertain behaviours associated with different stakeholders. Following the sequence of two-stage stochastic programming approach, decision variables for both the leader and the followers can be classified into design decisions that must be made “here-and-now” and operational decisions that are postponed to a “wait-and-see” mode after the realization of uncertainties (Gao and You, 2015b). Thus, in the first stage, both players will interact with each other to determine their optimal design strategies. Then, uncertainties associated with the leader and followers are realized. Based on the perceived uncertainty

knowledge as well as the predetermined design decisions, both types of stakeholders need to determine their optimal operational decisions. The uncertainties are depicted with discrete scenarios with given probabilities following the stochastic programming approach. The leader and the followers all strive to maximize their own expected net present values. The resulting problem can be formulated as a stochastic mixed-integer bi-level programming (MIBP) problem. Specifically, the upper-level problem is formulated as a mixed-integer nonlinear programming (MINLP) problem corresponding to the leader's optimization problem; the lower-level problems are formulated as linear programs corresponding to the optimization problems of the followers. By replacing the lower-level problems with their equivalent Karush-Kuhn-Tucker (KKT) conditions, we can reformulate this MIBP problem into a single-level mixed-integer nonlinear programming (MINLP) problem. To further improve the tractability of the resulting MINLP problem, the Glover's linearization approach is applied to reformulate it into an equivalent MILP problem (Glover, 1975). A large-scale case study of shale gas supply chains based on Marcellus Shale is presented to illustrate the applicability of the proposed modelling framework and solution strategy. Based on the optimization results, we can conclude that both players tend to choose more conservative strategies when considering the uncertainties in the non-cooperative supply chain optimization problem.

2. Problem statement

In this section, we formally state the problem of shale gas supply chain optimization considering uncertainties of EUR and market demand. The shale gas supply chain starts with a set of potential shale sites. Each shale site allows for drilling of multiple horizontal wells. The raw shale gas produced from shale sites, after pre-treatment near well head, is gathered and transported to midstream processing plants through pipelines. There is a set of potential processing plants to be constructed, where sales gas and natural gas liquids (NGLs) are separated and distributed to the market. Notably, in this shale gas supply chain, external shale sites and processing plants are considered, which can be regarded as recourse options to supplement the shale gas output or handle the extra shale gas exceeding the existing processing capacity. However, the utilization of such recourse options will lead to corresponding penalty costs. As mentioned above, two types of uncertainties are considered in this problem, namely the uncertain EUR of shale wells that can only be revealed after the drilling decisions are made, and the downstream uncertain demand of natural gas that will directly affect the overall economic performance of the whole supply chain. In this problem, the objective is to minimize the expected total cost throughout the shale gas supply chain based on the worst-case uncertainty distribution among all the distributions defined within the ambiguity set. Corresponding to the two-stage optimization structure of this DRO model, the first-stage decisions and the second-stage decisions summarized below.

First-stage design decisions:

- Development plan and drilling schedule for each shale site;
- Installation and capacity selection for the gathering pipeline network;
- Allocation and capacity design of processing plants.

Second-stage operational decisions:

- Amount of shale gas produced at each shale site in each time period;
- Amount of shale gas transported by each gathering pipeline from shale sites to processing plants;
- Amount of shale gas that exceeds the capacity of existing gathering pipeline;
- Amount of shale gas processed by existing and external processing plants in each time period;
- Amount of natural gas and NGLs to be distributed to the market;
- Amount of market demand satisfied by external shale gas suppliers.

3. Model Formulation and data-driven ambiguity set

The general form of a two-stage DRO model (P0) is presented as follows,

$$(P0) \quad \begin{aligned} \text{Objective: } & \min_{\mathbf{x}} \mathbf{c}^T \mathbf{x} + \sup_{P \in \mathcal{D}} E_P \{ l(\mathbf{x}, \xi) \} \\ \text{s.t. } & \mathbf{Ax} \leq \mathbf{b} \\ & l(\mathbf{x}, \xi) = \begin{cases} \min_{\mathbf{y}} \mathbf{d}^T \mathbf{y} \\ \text{s.t. } \mathbf{T}(\xi) \mathbf{x} + \mathbf{Wy} \leq \mathbf{h}(\xi) \end{cases} \end{aligned}$$

where \mathbf{x} denotes all the first-stage design decision variables that need to be determined before realization of uncertainties. \mathbf{y} denotes all the second-stage operational decision variables, which are dependent on the uncertain parameters ξ . The objective is to minimize the worst-case expected performance under all possible

uncertainty distributions \mathbf{P} in the ambiguity set \mathbf{D} . Corresponding to the two-stage optimization structure, the objective function comprises two parts, namely the first-stage cost determined by the first-stage design decisions and the worst-case expected cost associated with the second-stage operational decisions. As can be observed in the constraints of the second-stage subproblem, both the left-hand-side coefficient \mathbf{T} and the right-hand-side vector \mathbf{h} are influenced by uncertainties.

The ambiguity set \mathbf{D} includes a family of probability distributions with common statistical properties (Wiesemann et al., 2014):

$$\mathbf{D} = \left\{ \mathbf{P}_\xi \in \mathbf{M}_+ \left| \begin{array}{l} \mathbf{P}\{\xi \in \Xi\} = 1 \\ \mathbf{E}_P\{g_i(\xi)\} \leq \gamma_i, i = 1, 2, \dots, I \end{array} \right. \right\} \quad (1)$$

Such an ambiguity set can be expressed as the projection of an extended ambiguity set $\bar{\mathbf{D}}$ by introducing an I -dimension auxiliary random vector $\boldsymbol{\varphi}$, given as follows (Shang and You, 2018),

$$\bar{\mathbf{D}} = \left\{ \mathbf{Q}_{\xi, \boldsymbol{\varphi}} \in \mathbf{M}_+ \left| \begin{array}{l} \mathbf{P}\{(\xi, \boldsymbol{\varphi}) \in \bar{\Xi}\} = 1 \\ \mathbf{E}_Q\{\boldsymbol{\varphi}\} \leq \boldsymbol{\gamma} \end{array} \right. \right\} \quad (2)$$

where the domain of uncertainties is extended to a lifted support set $\bar{\Xi}$:

$$\bar{\Xi} = \left\{ (\xi, \boldsymbol{\varphi}) \left| \begin{array}{l} \xi \in \Xi \\ g_i(\xi) \leq \varphi_i, i = 1, 2, \dots, I \end{array} \right. \right\} \quad (3)$$

It has been proved that the ambiguity set \mathbf{D} is essentially tantamount to the set including all marginal distribution of ξ under $\mathbf{Q} \in \bar{\mathbf{D}}$ (Bertsimas et al., 2018).

Specifically, the following support set and moment functions are adopted in this study:

$$\Xi = \left\{ \xi \left| \xi_m^{\min} \leq \xi_m \leq \xi_m^{\max}, m = 1, 2, \dots, M \right. \right\} \quad (4)$$

$$g_i(\xi) = \max\{\mathbf{f}_i^T \xi - q_i, 0\}, i = 1, 2, \dots, I \quad (5)$$

To establish the lifted support set $\bar{\Xi}$, it requires us to determine \mathbf{f}_i and q_i associated with uncertainties of EUR and market demand. In this study, we adopt a systematic two-step data-driven procedure to determine \mathbf{f}_i and q_i based on the given data samples of shale well EUR and market demand (Shang and You, 2018). The basic idea of this two-step data-driven procedure is to first determine the projection directions $\{\mathbf{f}_i\}$ using the PCA approach, so that the EUR and market demand uncertainty data space is “decorrelated” along each direction, and the information overlap between projection directions is minimum. The following step is to set proper truncation points $\{q_i\}$ along each projection direction \mathbf{f}_i to capture the corresponding statistical information. To obtain a more tractable formulation of this two-stage DRO model as given in (P0), we adopt a pragmatic reformulation strategy based on affine decision rules to tackle this tractability issue (Ben-Tal et al., 2004). Notably, the resulting model formulation is a deterministic MILP, which can be solved efficiently by branch-and-cut algorithm implemented in MILP solvers such as CPLEX.

4. Application to a Shale Gas Supply Chain

In the case study of Marcellus shale gas supply chain, a total of five shale sites are considered. Each shale site allows for drilling of four to eight shale wells. The produced raw shale gas can be transported to three potential processing plants through gathering pipelines. Four capacity ranges are considered for the design of processing plants and gathering pipelines. A 10-year planning horizon consisting of 10 time periods is considered. As mentioned previously, two types of uncertainties are addressed, namely the EUR of shale wells and market demand of natural gas in each time period. The EUR data samples are generated based on the most up-to-date EUR data of 5,000 shale wells reported in the Marcellus shale play (Swindell, 2018). The market demand samples are estimated based on the natural gas consumption data provided by the U.S. Energy Information Administration (EIA, 2018b). The uncertainty involves 15 dimensions, corresponding to the EUR of five shale

sites and market demand in 10 time periods. For the proposed two-stage DRO model, a total of 1000 data samples are collected. We consider 5 truncation points along each principal direction, which result in 75 piecewise functions in the ambiguity set. The resulting MILP formulation of the DRO model has 122 integer variables, 164,731 continuous variables, and 78,519 constraints, and the global optimization of the resulting problem requires 1,101 CPU seconds. Due to the high-dimension EUR and market demand uncertainties involved in this case study, using traditional stochastic programming approach leads to computationally intractable optimization problems. Hereby, we only provide the benchmark performance considering a DRO model, a deterministic model, and a perfect information model in this section. The deterministic model is based on the nominal value of uncertain parameters, and the perfect information model is deterministic model that assumes actual realization of uncertainties is known. All the models are coded in GAMS 25.1.1. The MILP problems are solved using CPLEX 12.8. The absolute optimality tolerance is set to 10^{-6} .

To demonstrate the performance of the optimal design decisions obtained from different models, we simulate 100 data samples based on the same data source for benchmark purpose. The corresponding optimization results are summarized in Figure 1. The total costs in different scenarios associated with the DRO model, the deterministic model, and the perfect information model are presented in squares, triangles, and dots, respectively. With known realization of uncertainties, the perfect information model returns the minimum total cost that can be achieved in each individual scenario. The DRO model leads to a robust supply chain design that is able to maintain a very stable performance throughout the 100 scenarios. Although the deterministic model can outperform the DRO model in more than half of the scenarios, the optimal design decisions based on nominal uncertainty values are unable to adapt to the uncertain EUR and market demand in certain scenarios, thus resulting in extremely high penalty cost. The average total costs associated with the DRO model, the deterministic model, and the perfect information model based on the 100 scenarios are \$64.09 MM, \$73.44 MM, and \$19.36 MM, respectively.

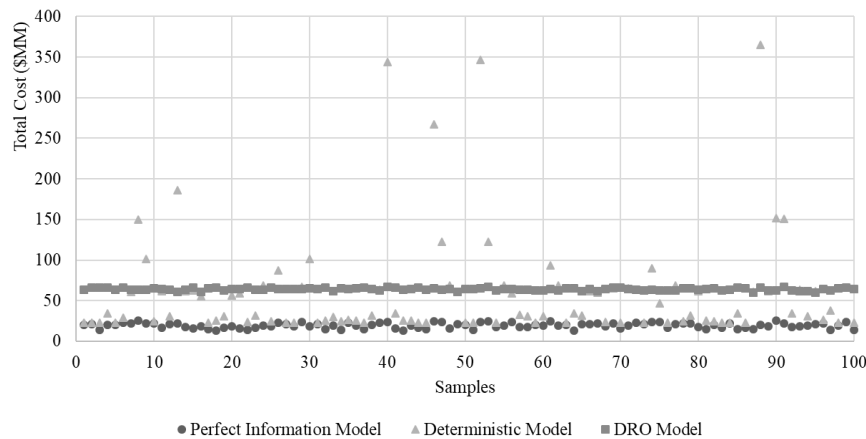


Figure 1: Performance comparison among the DRO model, the deterministic model, and the perfect information model based on simulation results of 100 scenarios

The detailed cost analysis based on the average performance of these 100 scenarios regarding these three types of models is provided in Figure 2. As can be observed, the robust design adopted in the DRO model results in a higher capital investment. The capital costs of processing plants and gathering pipelines associated with the DRO model are \$29.44 MM and \$0.52 MM, respectively. By contrast, the capital costs of processing plants and gathering pipelines associated with the deterministic model are \$13.43 MM and \$0.18 MM, respectively. In return, the supply chain design of the DRO model induces the least penalty cost, while the penalty cost of the deterministic model is much higher due to the ignorance of uncertainties. Moreover, since more shale wells are drilled in the optimal solution of the DRO model, the associated operating expenses regarding drilling, production, processing, and transportation are higher than those of the deterministic model.

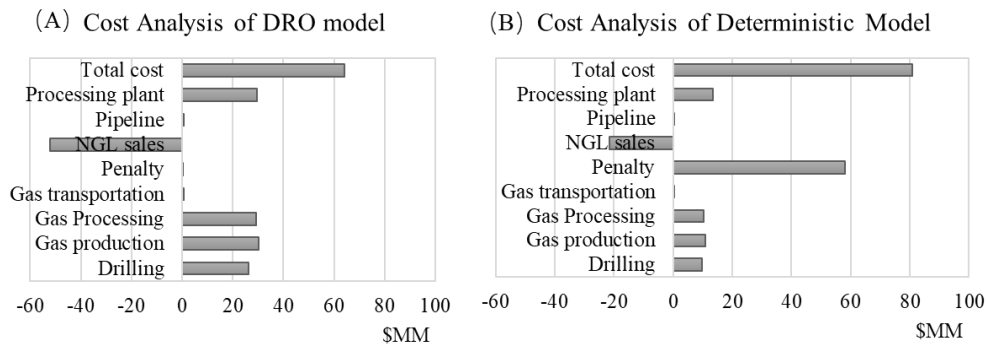


Figure 2: Cost analyses of (A) the DRO model and (B) the deterministic model.

In Figure 3, we summarize the optimal drilling schedules determined based on both the DRO model and the deterministic model. In the optimal solution of DRO model, all the five shale sites are developed with a total of 24 shale wells drilled. Meanwhile, only 9 shale wells located at shale site 2 and shale site 3 are drilled in the optimal solution of the deterministic model. With more shale sites developed, the optimal solution of the DRO model can effectively hedge against the EUR uncertainty of different shale sites. On the other hand, more shale wells produce more shale gas, which is sufficient to satisfy the uncertain market demand throughout the planning horizon. Besides the different drilling strategies, a similar drilling pattern can be observed in both models, where more shale wells are drilled in the first few years, and extra shale wells are drilled later to compensate the decreasing production of existing shale wells.

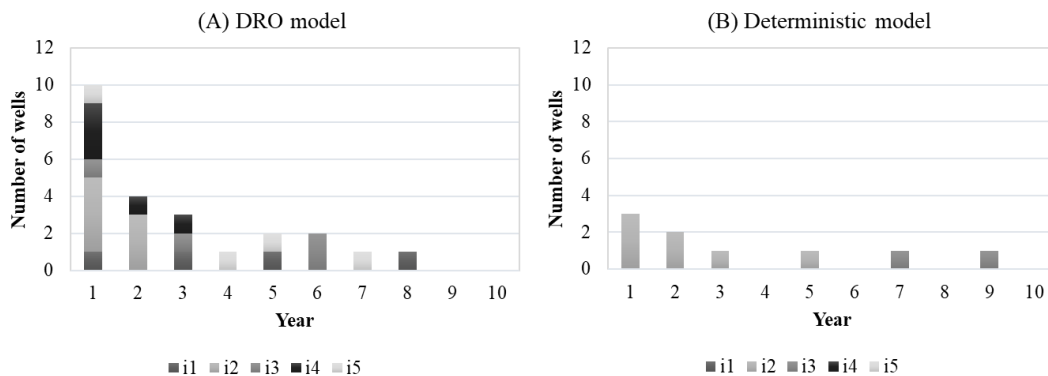


Figure 3: Optimal drilling schedules of (A) the DRO model and (B) the deterministic model.

Next, we investigate the actual working conditions of designed infrastructure in their corresponding shale gas supply chains in Figure 4. The shale gas production from different shale sites in each time period is presented with stacked columns. Due to the EUR uncertainty, the corresponding error bars are provided to describe the possible fluctuation of actual shale gas output. The designed working capacity of the processing plant is depicted by a green dashed line. From this figure, we can clearly identify the advantage of the proposed DRO model in leading to more robust supply chain designs that are capable of hedging against the EUR uncertainty nearly perfectly. The designed processing plant is sufficient to handle the highest possible shale gas output without wasting any processing capacity. By contrast, in the deterministic model, the designed processing plant is only capable of handling the nominal shale gas production. When EUR uncertainty is involved, there is a high chance that extra shale gas processing capacity is required. Similarly, by looking into the working conditions of gathering pipelines based on the optimal designs obtained from the DRO model and the deterministic model. We can identify that the gathering pipelines designed based on the DRO model hold a robust performance to address the uncertain shale gas production from each shale site. In the deterministic model, however, the designed pipelines lack the potential to transport the uncertain shale gas output from corresponding shale sites.

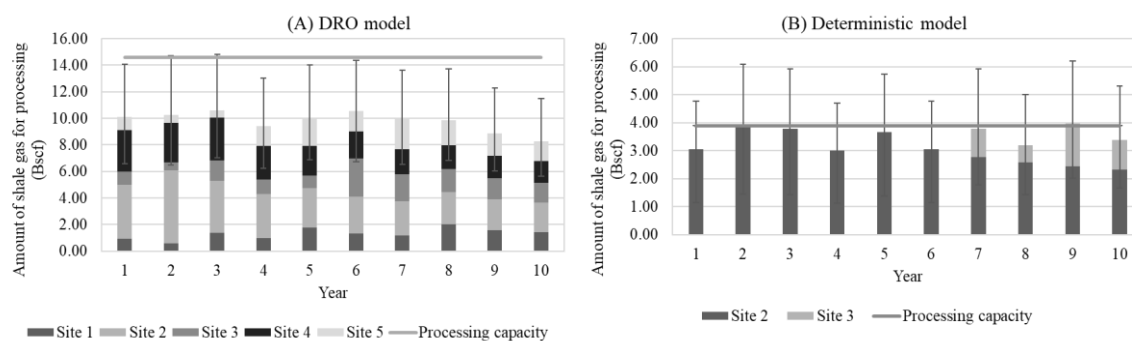


Figure 4: Working conditions of processing plants based on the optimal designs obtained from (A) the DRO model and (B) the deterministic model

Finally, when we focus on the downstream of this shale gas supply chain. We can observe that a relatively conservative shale gas production strategy is adopted in the DRO model, such that even the lowest total shale gas production is sufficient to satisfy the highest possible market demand. Thanks to the robust development strategy in the optimal solution to the DRO model, no external shale gas is needed to satisfy the market demand. Meanwhile, since the impact of uncertain EUR and market demand is not considered in the deterministic model, only two shale sites are developed, and the resulting total shale gas production is much less than that of the DRO model. Consequently, the uncertain market demand cannot be satisfied solely by the shale gas production from the developed shale sites, and the external shale gas supply is required throughout the planning horizon. As can be expected, the heavy reliance on external supplier can render the shale gas supply chain fragile and less competitive.

5. Conclusions

In this study, we proposed a data-driven two-stage DRO model to address the optimal design and operations of shale gas supply chains under the uncertainties of EUR and market demand. A two-step data-driven procedure based on PCA and first-order deviation functions, as well as industry data, was introduced to construct the ambiguity set of the DRO model. We further presented a tractable MILP reformulation by taking advantage of affine decision rules and duality theorem. To illustrate the applicability of proposed modelling framework and solution approach, a case study of Marcellus shale gas supply chain was presented. From the optimization results, we concluded that the proposed DRO model can provide robust supply chain designs that can reliably maintain a competitive performance under uncertainty.

References

- Ben-Tal A., Goryashko A., Guslitzer E., Nemirovski A., 2004, Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99, 351-376.
- Bertsimas D., Doan X.V., Natarajan K., Teo C.-P., 2010, Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35, 580-602.
- Bertsimas D., Sim M., Zhang M., 2018, Adaptive distributionally robust optimization. *Management Science*, DOI: 10.1287/mnsc.2017.2952.
- Delage E., Ye Y., 2010, Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58, 595-612.
- EIA, 2018a, Annual energy outlook 2018. U. S. Energy Information Administration, Washington, DC 20585, USA, <<https://www.eia.gov/outlooks/aeo/pdf/AEO2018.pdf>>, accessed 03/04/2019.
- EIA, 2018b, Natural gas consumption by end use. <https://www.eia.gov/dnav/ng/ng_cons_sum_dc_u_nus_a.htm>, accessed 03/04/2019.
- Gao J., You F., 2017, Design and optimization of shale gas energy systems: Overview, research challenges, and future directions. *Computers & Chemical Engineering*, 106, 699-718.
- Shang C., You F., 2018, Distributionally robust optimization for planning and scheduling under uncertainty. *Computers & Chemical Engineering*, 110, 53-68.
- Swindell G.S., 2018, Estimated ultimate recovery (EUR) study of 5,000 Marcellus shale wells in Pennsylvania, <http://www.gswindell.com/marcellus_eur_study.pdf>, accessed 03/04/2019.
- Wiesemann W., Kuhn D., Sim M., 2014, Distributionally robust convex optimization. *Operations Research*, 62, 1358-1376.