# Surface Facility Optimization for Combined Shale Oil and Gas Development Strategies

*Agustín F. Montagna[a,b], Diego C. Cafaro[a,b], Ignacio Grossmann[c], Ozgur Ozen[d], Yufen Shao[e], Ti Zhang[f], Yuanyuan Guo[e], Xiao-Hui Wu[g] & Kevin C. Furman[e]*

[a] *INTEC(UNL-CONICET), Güemes 3450, 3000 Santa Fe, ARGENTINA*
[b] *Fac. de Ing. Química, Universidad Nacional del Litoral, Santiago del Estero 2829, 3000 Santa Fe, ARGENTINA*
[c] *Dept. of Chem. Eng., Carnegie Mellon University, 5000 Forbes Ave, 15213 Pittsburgh, PA, US*
[d] *ExxonMobil Global Projects Co, 22777 Springwoods Village Parkway, 77389 Spring, TX, US*
[e] *ExxonMobil Upstream Research Co, 22777 Springwoods Village Parkway, 77389 Spring, TX, US*
[f] *(Previously) ExxonMobil Upstream Research Co, 22777 Springwoods Village Parkway, 77389 Spring, TX, US*
[g] *ExxonMobil Upstream Integrated Solutions Co, 22777 Springwoods Village Parkway, 77389 Spring, TX,US*

## *Abstract*

In the context of a global energy transition, oil and gas will remain an important part of the energy mix, especially in developing countries. The challenge of energy companies is to adapt to a changing policy and investment landscape, and still remain competitive. In this work we present a generalized optimization framework for the design of oil and gas gathering networks accounting for combined shale oil and gas development strategies. We develop mixed-integer linear (MILP) and quadratically constrained models (MIQCP) to optimally determine the network of pipelines, separation, processing and delivery facilities for both oil and gas. In contrast to previous approaches, the networks are built with no predetermined number of echelons. We assume that there is a set of generic nodes to be connected among themselves to reach the final destinations. By including pressures as decisions variables, flowrates and flow directions can be optimally handled along the time horizon to make a better use of the transportation capacity. Stochastic programming extensions of the models permit to determine the network of surface facilities that maximizes the expected net present value of the project under uncertain scenarios. Oil and gas prices may significantly change in the future, being not clear if the company focus will be on developing wells producing more oil than gas, or vice-versa. Shale oil and shale gas wells usually coexist in nearby regions of the same formation, thus being a nontrivial decision where and when to build and expand the gathering networks. We assess the potential of the formulations by solving three illustrative case studies and a real-world problem from the unconventionals' industry.

## *Introduction*

The energy industry is facing the challenge to simultaneously meet decarbonization goals while still fulfilling the expected demands for oil and gas. Under an ongoing global energy transition, oil and gas remain an important part of the energy mix, especially in developing countries (IEA, 2021). Even the most aggressive decarbonization forecasts predict a long-term role for oil and gas although demand levels are expected to decrease steadily under sustainable development scenarios (e.g. net zero emissions by 2050). In the United States, India, and China, the three largest greenhouse gas emitters, natural gas in particular has the potential to lead the energy transition to low carbon economy for decades, depending on the policy mechanisms and technologies in place (Johnston et al., 2020). The challenge for the oil and gas industry is to both engage and adapt to a changing policy and investment landscape, and still remain competitive.

In this context, the optimization of the network of facilities to gather, process and deliver unconventional oil and gas production has received increasing attention from the research community in the las 10 years. There are two main reasons that motivated the interest in this problem: (1) the rapid expansion of the unconventionals' industry, particularly in the United States (EIA, 2021); and (2) the steep decline that characterizes the productivity of unconventional wells, needing to make decisions more accurately and more often. Energy companies have made efforts to properly plan their operations across shale oil and gas formations in order to stay profitable. Significant advances in horizontal drilling and hydraulic fracturing technologies have facilitated the exploitation of unconventional resources, but economic margins are usually low due to the heavy investments that are required. As a result, many research groups around the globe have developed sophisticated mathematical programming techniques to leverage the decision making process (Gao and You, 2015).

***Previous Contributions***

The optimal design of gathering networks for unconventional fuels production has been first addressed by Cafaro and Grossmann (2014). In their seminal work, they propose a non-convex mixed integer non-linear programming formulation (MINLP) to optimize the planning of drilling operations over a shale gas area, while simultaneously determining the optimal location and size of compressors, pipelines and gas processing plants. Drouven and Grossmann (2016) expand the scope of the problem by including gas quality variations, differentiating drilling and fracturing schedules, and discretizing economies of scale, particularly for pipeline sizing. Tan and Burton (2016) develop a two-stage stochastic programming formulation for the allocation of small-scale mobile plants for gas processing. A real world problem from the Bakken shale play is solved by means of a mixed-integer linear programming (MILP) formulation. The aim is to optimally acquire and allocate flexible gas-to-liquids and liquefaction plants for the separation of natural gas liquids (NGL) and the production of liquefied natural gas (LNG), respectively. In contrast to the previous models, the pipeline sizing is out of the scope of this work.

The same authors (Tan and Burton, 2017) propose a similar approach for a country-wide supply chain optimization, comprising multiple plays (Bakken, Utica, Marcellus, Niobrara, Permian, Haynesville and Eagle Ford). They seek to optimize the location and capacity of chemical plants (hydroskimming refineries, gas-to-liquids and liquefaction plants) to convert shale oil and gas into LNG, gasoline, kerosene, diesel and residual fuel oil. In this problem, they also address the selection of transportation modes (pipeline, road, rail or barge) to carry the fuels across the network. Uncertainty in the optimal design and operation of shale gas supply chains has been also addressed by Gao et al. (2019) by means of 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 all of these approaches the objective is the maximization of the net present value (NPV) over a time horizon that comprises 10 to 20 years and is usually discretized into annual quarters.

Allen et al. (2019) also aim to optimize the use of modular and transportable plants. Gas processing facilities can be composed of multiple modular plants operating in parallel. The modular plants can be reallocated within the field to other processing facilities to face the uncertainty in production that comes with developing a shale gas field. The authors propose to formulate the optimization problem as a two-stage multi-period or a multi-stage stochastic program, depending on the structure of the uncertain production forecasts. Based on similar models, Hong et al. (2020a) develop a two-stage stochastic formulation for determining the optimal start times for shale gas wells production, also deciding on the acquisition, allocation and mobilization of modular gas processing units (GPU). The model accounts for capacity selection, installment and mobilization planning, as well as salvage operations for the GPU.

Through a more detailed MINLP model, Hong et al. (2020b) address the optimal design of a gathering pipeline system for shale gas production. To solve the problem the authors develop a piece-wise linear approximation based on discrete ranges for the flowrates, while the set of potential connections (3D pipeline layout) is predefined by first solving an ant-colony algorithm. The problem is solved for a time horizon comprising 3 periods, over a limited geographical area. Kröetz et al. (2019) develop a multiobjective integer-programming model for pipeline siting that incorporates habitat externalities to estimate the trade-offs between pipeline development costs and habitat impacts. More recently, Montagna et al. (2021) present a comprehensive MINLP formulation for the optimal design of the pipeline network connecting shale oil wells to tank batteries for the separation of gas and water from the production stream. The model accounts for detailed multiphase pressure drop calculations to size the pipeline diameters according to the product flow to handle over time. Like most contributions addressing the optimal design of pipeline networks, reference values for inlet and outlet pressures at the segments are given beforehand to simplify the pipeline sizing equations (Cafaro and Grossmann, 2014; Drouven and Grossmann, 2016; Montagna et al., 2021).

Most of previous formulations assume that facilities and pipeline interconnections should configure a network with a fixed number of echelons. More specifically, there is a set of production nodes to be connected with one or several junction nodes, which in turn are connected to one or more separation and processing nodes, to finally reach the markets. The shale gas supply chain network presented by Cafaro and Grossmann (2014) is a typical example. It is interesting to note that the product state evolves with each stage. In the shale gas network, the raw gas produced in the wellpads is dehydrated and compressed in junction nodes, cryogenically separated in distillation plants, and the components (natural gas, ethane and LPG) are finally delivered to demand nodes (connections to midstream distributors). A similar configuration has been adopted by Guerra et al. (2016), also integrating the water management, and more recently by Montagna et al. (2021) for the optimal design of gathering networks collecting oil, gas and water from the wellpads over a shale formation. In the latter case, the main challenge is to manage multiphase flows along pipelines until reaching the tank batteries for separation.

In contrast to previous contributions, this work presents a generalized framework in which the oil and gas gathering networks are built with no predetermined number of echelons. We assume that there is a set of generic nodes to be connected among themselves to reach the final destinations. In any of these nodes, facilities for merging, splitting, storing, separating, processing and/or delivering flows will be installed to make the oil and gas flows be ready for use. One of the major differences with regards to the models with a fixed number of echelons is that the flow direction may be reversed in any pipeline segment over the time horizon. Besides that, given that the number of segments connecting a source node to a delivery node is optimally determined by the model an additional challenge is to track pressures along the paths. Based on the resulting network design one should accurately define the inlet and outlet pressures at every segment for every time period. By including pressures as decisions variables to be determined, flowrates and flow directions can be optimally handled along the time horizon to make a better use of the pipeline transportation capacity. However, nonlinear constraints are required, thus posing further challenges on the modelling and computational sides.

Finally, this work also aims to address one of the big questions faced by facility planners: how to adapt to changing policies and development plans. By means of stochastic programming extensions of the mathematical formulations, we seek to determine the optimal network of oil and gas surface facilities that is able to maximize the expected net present value of the project under uncertain scenarios. We assume that oil and gas prices may significantly change in the future, and from this it is not clear if the company focus will be on developing wells producing more oil than gas (i.e., with low gas-to-oil or GOR ratios), or vice-versa. The interesting feature is that both types of wells (with either low or high GOR) usually coexist in nearby regions of the same formation, thus being a nontrivial decision where and when to build and expand the gathering networks.

In the next section we formally define the problem of designing facility networks under combined shale oil and gas development strategies, and then introduce the model assumptions and their fundamentals. We develop MILP and MIQCP (mixed integer quadratically constrained) models to optimize oil and gas gathering networks, and afterwards present an integrated framework also accounting for uncertainty. We finally solve three illustrative case studies and a real-world problem from the unconventionals' industry to discuss the potential of the proposed optimization approaches and draw conclusions.

## *Problem Definition*

The optimization problem addressed in this work can be stated as follows. Given:

(a) A hybrid shale oil and shale gas exploitation region where dozens of multiwell pads are to be developed in a long term planning horizon. The geographical location of the wellpads and the number of wells per pad are also given.
(b) A finite set of alternative development plans based on different oil and gas price scenarios. They are previously defined by company's policy-makers and can be roughly categorized in three groups: (i) shale-oil focused, (ii) shale-gas focused, (iii) mixed strategies. The pace-of-activity (average number of wells completed per year) is also dependent on the scenario. For each alternative plan, drilling and completion dates are given, together with the expected productivity profiles from the wellpads over the time horizon.

The aim is to design an integrated network of surface facilities to gather flows, separate phases, process product streams and deliver both shale oil and shale gas production over the time horizon, while maximizing the expected net present value of the project under different development strategies.

Unconventional wells produce three basic components: oil, gas and water, which need to be processed in properly sized and located facilities. Both shale oil and shale gas wells may be drilled and fractured over the same formation, yielding different Gas-to-Oil Ratios (GOR). GOR is the critical measure to distinguish between oil, wet gas and dry gas wells, and it is strongly dependent on the geographical location of the wellpads. Shale oil wells usually feature low GOR, while shale gas wells mainly produce natural gas, together with small fractions of oil and condensates, yielding high GOR. Figure 1 illustrates oil and gas windows over the Permian basin in the US and the Vaca Muerta formation in Argentina.

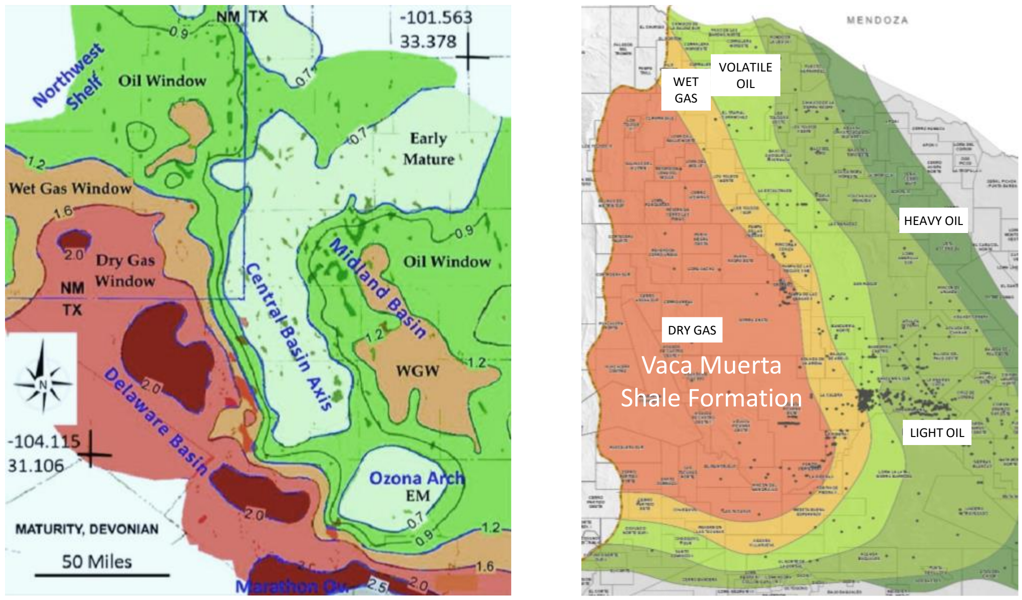


*Figure 1. Shale oil and shale gas windows over the Permian Basin (US) and the Vaca Muerta Formation (Argentina)*

Finally, time varying Water-to-Oil Ratios (WOR) are commonly seen in both shale oil and shale gas wells. In this problem, the production of oil, gas and water from every single wellpad over the time horizon is given data, according to the development plan of each price scenario.

On the other hand, the network of surface facilities basically comprises:

1. Two-phase separators, primarily separating natural gas from the liquids (oil+water).
2. Three-phase separators (free water knockouts, heater treaters, gas scrubbers, among others).
3. Gas dehydration systems.
4. Gas compressors and pumps.
5. Pipelines, which can be classified into flowlines (connecting wells to primary separators), liquid pipelines (conveying oil, water, and/or oil emulsion), low pressure gas pipelines and high pressure gas pipelines.
6. Valves, control devices and auxiliary equipment.

All these elements are usually integrated into units or facilities of different sizes and capacities, aiming at processing the wells' production in a sequence of steps, adding value to every single component to finally yield sellable products. Among these units, we may find (i) primary two-phase separation units, (ii) oil tank batteries, (iii) gas production units (GPU), (iv) centralized oil and/or gas processing facilities, and (v) storage and delivery points. Wellpads need to be connected to separation and/or processing units through flowlines, and processing units need to be further connected among themselves through liquid and/or gas pipelines to finally deliver the product flows to the market.

The problem is aimed at optimally determining the number, location and size of processing facilities, together with the network of pipeline connections (pipeline diameters and lengths) to gather, process and deliver shale oil and shale gas flows from the unconventional formation. The time for facility investments and expansions is critical. Although recourse decisions might be made as the development plan becomes clearer, the starting network of facilities (planned for the first time period) should be ready to face any scenario effectively. Economies of scale also play a key role. The goal is to make these decisions so as to maximize the Expected Net Present Value (ENPV) of the project accounting for a finite set of oil and natural gas price scenarios, for each of which an alternative development plan is defined by the company.

## *Model Assumptions*

1. The development of shale oil and shale gas wells is organized in rows of wellpads (see Figure 2). This arrangement maximizes the recovery of resources from the shale formation by intensively drilling and fracturing horizontal wells in a compact area, also minimizing resource mobilization (Ondeck et al., 2019).
2. There is a finite set of alternative processing facilities, with different configurations and capacities, according to their purpose. Some of them may be basic two-phase separators (horizontal or vertical vessels) to primarily split the liquids from the raw gas when the flow exits the wellheads; while others may be integrated facilities featuring complex configurations meant to dewater, purify and desalt the oil, and/or conveniently dry the natural gas, before deliver them to midstream distributors.
3. We assume that liquid flows (a mixture of oil, free water and emulsions) are separated from raw gas at every single row. However, the number, size and time for installing these two-phase separators are model decisions.
4. Each flow (liquid/gas) leaving a wellpad reaches the corresponding collecting node of the row through a trunk flowline. The location of the collecting nodes along the row is given data.

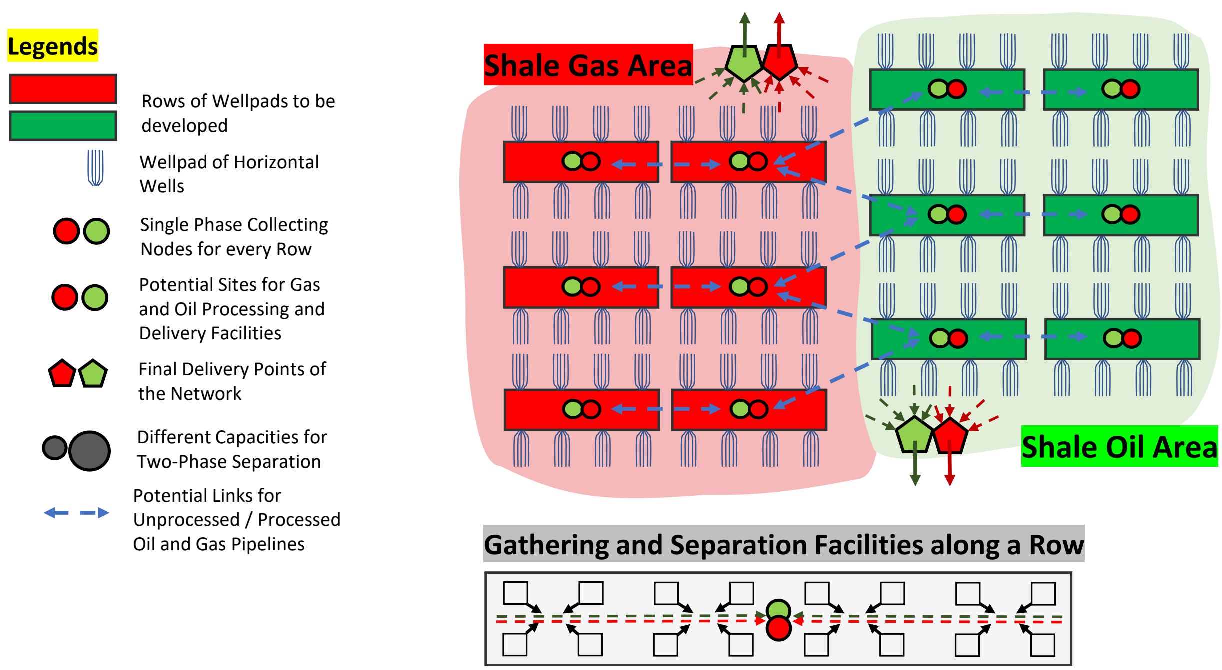


*Figure 2. Rows of wellpads and superstructure of alternatives to gather oil and gas flows.*

5. Both liquid and gas flows have to be sent to two successive stages after primary separation: (1) processing, (2) storage and delivery. If no processing facility has been installed in the row, the production flow reaching the collecting node needs to be derived to an adjacent row. An illustration of possible locations for oil and gas processing facilities is given in Figure 3.
6. After processing, oil and gas flows also need to reach a delivery node. The flows need to subsequently move among rows until reaching a delivery facility installed in a certain row, with enough capacity. Possible locations for oil and gas delivery nodes are also given in Figure 3.
7. The pressure of shale gas flows can only be boosted at the wellpads, at the processing facilities and at the delivery nodes. Building compressor stations at intermediate nodes is not allowed.
8. A finite number of alternative development strategies for shale oil and/or shale gas wells are given beforehand. Some of them may be focused in the shale oil extraction, others in the shale gas production, and there may be also combined strategies by simultaneously assigning rigs and fracturing crews to both shale oil and shale gas regions. Each of these plans includes:
    a. Geographical location of the wellpads.
    b. Number of wells to develop in each wellpad.
    c. Drilling and completion dates of the wells in the pad.
    d. Productivity of oil, gas and water for every wellpad over the time horizon.
    e. Oil and gas price forecasts for each scenario.
    f. Probability of following each strategy according to the oil and gas price forecast.

    We assume that each well development plan has been previously determined according to the corresponding oil and gas price forecast scenario. The number of active rigs determines the pace-of-activity in both shale oil and shale gas regions, and will be usually higher under optimistic scenarios, predicting higher oil and gas prices.

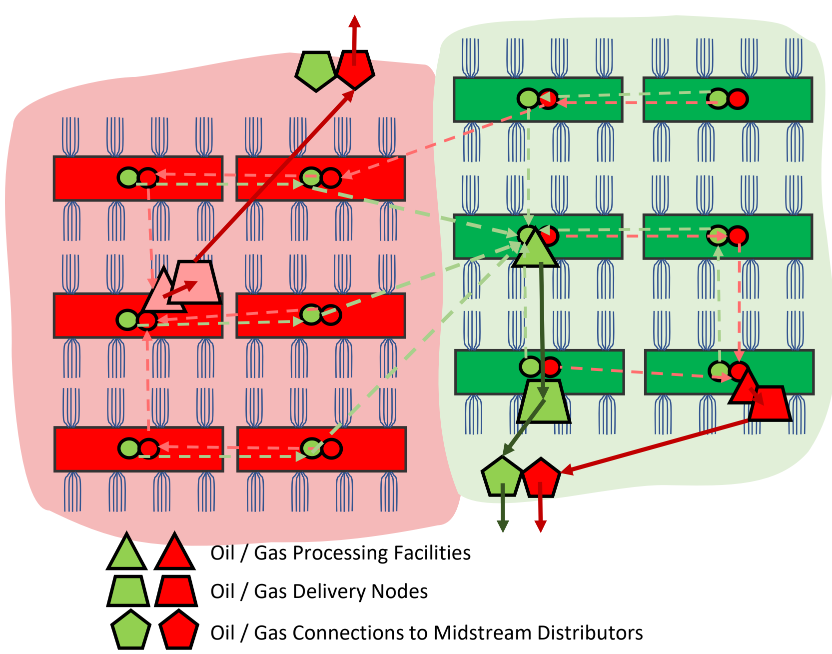


*Figure 3. A network of processing facilities, delivery nodes and pipelines. Dotted lines represent pipelines for unprocessed oil and gas flows.*

## *Mathematical Formulation*

To solve this problem, we propose a multiperiod, superstructure-based mathematical formulation that is presented in four parts: (a) a mixed-integer linear programming (MILP) model for the shale oil gathering network design, (b) a mixed-integer, quadratically constrained (MIQCP) model for the shale gas network design, (c) an MIQCP model for the integrated design of shale oil and shale gas gathering networks, and (d) a multi-stage stochastic programming model, for combined shale oil and gas development strategies.

The first two formulations are closely related to recent contributions to the optimal design of shale oil and shale gas gathering networks (Montagna et al., 2021; Drouven & Grossmann, 2016; Cafaro & Grossmann, 2014). However, there are two new challenges faced in this work: (i) the number of echelons in the gathering network (i.e., the "steps" or pipeline segments in the path from the wells to the delivery nodes) is not given beforehand; and (ii) the flow direction in oil and gas pipelines can be reversed in different time periods. These features are particularly challenging in the shale gas network due to the need to conveniently set the gas pressures at every node during the time horizon. Finally, the third comprehensive model integrates the elements of the previous formulations into a single optimization approach under uncertain development scenarios.

### *MILP Model for Shale Oil Gathering Network Design*

The model is divided in groups of equations, according to their aim. The first group deals with the first step in the gathering network: moving unprocessed (raw) flows from the rows of wellpads to the processing facilities. The second group accounts for the second step: moving processed flows to delivery nodes. Finally, the objective function is presented and described in the third subsection. We note that equations are defined for a generic product *k*, since most of them will be also used to model the shale gas gathering network.

*Gathering Unprocessed Flows*

The network design is based on the the concept of rows of wellpads. The parameter $p_{k,w,r,t}$ represents the amount of component *k* produced by wellpad $w \in W_r$ along row *r* during the time period *t.* All product flows need to be sent to processing. In eq. (1), the continuous variable $P_{k,r,t}$represents the total amount of *k* that is processed in the same row *r*, only if a processing facility has been already installed in that row. If there is no processing facility in *r* or its capacity is no large enough, the unprocessed flow should be sent to an adjacent row $r' \in J_r$. The total amount of component *k* sent from *r* to *r'* during *t* is given by $Q_{k,r,r',t}$. Note that production flows accumulate in their way to the row with a processing facility for component *k*.

$$\sum_{w \in W_r} p_{k,w,r,t} + \sum_{r' \in J_r} Q_{k,r',r,t} = P_{k,r,t} + \sum_{r' \in J_r} Q_{k,r,r',t} \qquad \forall k,r,t \tag{1}$$

The overall capacity of the processing facilities already installed in row *r* is imposed as an upper bound for $P_{k,r,t}$ by eq. (2). The binary variable $x_{k,s,r,t'}$ stands for the decision of installing a new facility of size *s* (whose processing capacity for component *k* is given by $pc_{k,s}$) in row *r* at time period *t'*. By building new facilities in the same location the model increases the processing capacity of that node. For simplicity, we assume that every new facility is ready for utilization in the same period that it is installed. However, installation lead-time given as a number of time periods can be handled in a straightforward manner. Also note that just a subset of periods $t' \in TI \subseteq T$ is usually proposed for new investments.

$$P_{k,r,t} \leq \sum_{\substack{t' \in TI \\ t' \leq t}} \sum_{s} pc_{k,s}\, x_{k,s,r,t'} \qquad \forall k, r, t \tag{2}$$

Analogously, the transportation capacity between rows *r* and *r′* to move unprocessed component *k* is also limited by the diameter of the pipelines connecting both rows, as imposed by eq (3). We assume that there exists a set of alternative pipeline diameters *d* that will be conveniently selected by the model through the binary variable $u_{k,d,r,r',t'}$. Similar to the model of Cafaro and Grossmann (2020) for water network design, both directions *r-r′* and *r′-r* are comprised in the same equation, although only one of the terms will be active in the optimum to reduce operating costs. Note that this equation is proposed for incompressible flow transportation, and will be revisited in the upcoming section. We refer the reader to the work by Cafaro and Grossmann (2020) for more details on the determination of the transportation capacity and pumping costs along incompressible flow pipeline segments.

$$Q_{k,r,r',t} + Q_{k,r',r,t} \leq \sum_{\substack{t' \in TI \\ t' \leq t}} \sum_{d} tc_{k,d}\, u_{k,d,r,r',t'} \qquad \forall k = oil, r, r' \in J_r, r < r', t \tag{3}$$

*Delivering Processed Flows*

The total amount of component *k* leaving a processing facility ($\alpha\, P_{k,r,t}$) needs to be sent to a delivery node. For simplicity, we assume a fixed fraction $\alpha \leq 1$ as the conversion factor from unprocessed to processed volumes. In the case of shale oil, this factor deducts the remaining traces of water, dissolved gas and solids contained in the flow after primary separation at the wellpads. The continuous variable $D_{k,r,t}$ represents the total amount of *k* that is delivered to midstream distributors from the same row *r* (a delivery facility has to be operative in that row) while $R_{k,r,r',t}$ is the flow of processed component *k* sent from *r* to *r′* before reaching a delivery node. Eqs. (4), (5) and (6) are analogous to (1), (2) and (3). However, instead of selecting the capacity of the delivery node from a set of alternatives, we assume that facilities has to built and enlarged following subsequent expansions, as captured by eq. (7). The binary variable $y_{k,e,r,t'}$ equals one if the *e*-th expansion (being *e* = 1 the initial installation) of the delivery node in row *r* is ready for operation in period *t′*.

$$\alpha\, P_{k,r,t} + \sum_{r' \in J_r} R_{k,r',r,t} = D_{k,r,t} + \sum_{r' \in J_r} R_{k,r,r',t} \qquad \forall k, r, t \tag{4}$$

$$D_{k,r,t} \leq \sum_{\substack{t' \in TI \\ t' \leq t}} \sum_{e} dc_{k,e}\, y_{k,e,r,t'} \qquad \forall k, r, t \tag{5}$$

$$R_{k,r,r',t} + R_{k,r',r,t} \leq \sum_{\substack{t' \in TI \\ t' \leq t}} \sum_{d} tc_{k,d}\, v_{k,d,r,r',t'} \qquad \forall k = oil, r, r' \in J_r, r < r', t \tag{6}$$

$$y_{k,e,r,t} \leq \sum_{\substack{t' \in TI \\ t' < t}} y_{k,e-1,r,t'} \qquad \forall k, r, t \in TI \subseteq T \tag{7}$$

*Objective Function: Minimize the Net Present Costs of the Facilities*

If production forecasts are given, incomes from product sales can be estimated beforehand. Indeed, facility planners simply seek to minimize the net present cost (NPC) of the facilities (including pipelines) that are required to gather, process and deliver the flows over the time horizon. Such objective function is presented in eq. (8), where *i* is the interest rate to discount cashflows back to present; while *ipf*, *idp* and *ipl* are unit investment costs for processing facilities, delivery nodes and pipelines of different capacities or diameters, respectively. In turn, *opf*, *odp* and *opl* are unit operating costs for processing, delivering and moving unprocessed/processed flows through pipelines. For simplicity, unit operating costs are considered independent of the capacity of the facilities, but they may vary according to their location. Similarly, unit transportation costs are assumed to be independent from the pipeline diameter, but they vary with the pipeline length (distance from *r* to *r'*) and flow direction (see Cafaro and Grossmann (2020) for more details).

$$Min\ NPC = \sum_t \frac{1}{(1+i)^{t-1}} \left\{ \begin{array}{c} \sum_{k,s,r} ipf_{k,s}\ x_{k,s,r,t} + \sum_{k,e,r} idp_{k,e}\ y_{k,e,r,t} + \sum_{\substack{k,d,r \\ r' \in J_r \\ r' > r}} ipl^U{}_{k,d,r,r'}\ u_{k,d,r,r',t} + ipl^P{}_{k,d,r,r'}\ v_{k,d,r,r',t} \\ \sum_{k,r} opf_{k,r}\ P_{k,r,t} + \sum_{k,r} odp_{k,r}\ D_{k,r,t} + \sum_{\substack{k,r \\ r' \in J_r}} opl^U{}_{k,r,r'}\ Q_{k,r,r',t} + opl^P{}_{k,r,r'}\ R_{k,r,r',t} \end{array} \right\} \quad (8)$$

In summary, the MILP model for the optimal design of the oil gathering network seeks to minimize function (8), subject to constraints (1) to (7).

## ***MIQCP Model for Shale Gas Gathering Network Design***

To address the shale gas gathering network, the main challenge is the need to track pressures at the rows connections. In contrast to liquid pipelines, for which the maximum flow rate is usually assumed to be proportional to the cross sectional area (Cafaro & Grossmann, 2020), the pipeline transportation capacity for compressible fluids can be modified by managing pressures. In our problem, the number of segments along which the flow of unprocessed/processed gas moves to reach a processing/delivery node depends on the network design. Moreover, compressor stations are only available in processing and delivery nodes (assumption 7), thus requiring an accurate setting of inlet and outlet pressures at every segment. The aim of the model presented in this section is twofold: (a) to optimally design the network of pipelines to gather the shale gas produced at the rows, by adopting the most convenient diameter and time to build each connection, and (b) to handle pressures, and therefore shale gas flowrates and flow directions, along the time horizon.

*Weymouth Correlation for Shale Gas Hydraulics*

According to the SPE (SPE, 2021), the Weymouth correlation (Weymouth, 1912) is ideal for designing pipelines in gas field gathering systems. A simplified form of the correlation is presented in eq. (9), for a given pipeline length and temperature.

$$F = \frac{1.1\ d^{2.667}}{L\ S\ Z\ T_1} ({P_1}^2 - {P_2}^2)^{0.5} \ \leftrightarrow \ F^2 = \varphi^2\ d^{5.334} ({P_1}^2 - {P_2}^2) \quad (9)$$

*F* is the gas flow rate in $10^6$scf/day, *d* and *L* are the pipeline inside diameter (in inches) and length (in feet), *S* is the specific gravity of the gas in normal conditions (relative to air), *Z* is the gas compressibility factor, $T_1$ is the temperature of the gas inlet (in °R) while $P_1$ and $P_2$ are the inlet and outlet absolute pressures (in psi). The parameter $\varphi$ synthetizes all the factors that are assumed to be constant for a given pipeline segment and is also used for unit conversion. Based on this equation, we introduce the variable $PU_{r,t}^{sq}$ to account for the square pressure at the junction of row *r* during period *t*, for unprocessed shale gas transportation. An analogous variable ($PP_{r,t}^{sq}$) is used for processed gas flows.

*Gas Pipeline Transportation Capacity*

By eq. (10) the difference of square pressures between two adjacent rows *r* and *r'* determines the maximum admissible flow rate through a pipeline of diameter *d* connecting *r* to *r'* ($MaxFlowU_{r,r',d,t}$). Note that because pipeline flows can be reverted it is necessary to enforce the flow to be zero when the difference of square pressures is negative (the gas moves in the opposite direction). This is imposed by eqs. (11) and (12), where $u_{dir_{r,r',t}}$ is a binary variable that takes value one if the unprocessed gas flows from *r* to *r'* during time period *t*, and zero otherwise.

$$MaxFlowU_{r,r',d,t} \leq \varphi \;\; diam_d^{2.667} \left(\Delta PU_{r,r',t}^{sq}\right)^{0.5} \qquad \forall r,r' \in J_r, d, t \tag{10}$$

$$u_{dir_{r,r',t}} + u_{dir_{r',r,t}} = 1 \qquad \forall r,r' \in J_r, t \tag{11}$$

$$\begin{gathered} \Delta PU_{r,r',t}^{sq} \leq (PU_{r,t}^{sq} - PU_{r',t}^{sq}) + \Delta spu_{r,r'}^{Max}(1 - u_{dir_{r,r',t}}) \\ \Delta PU_{r,r',t}^{sq} \leq \Delta spu_{r,r'}^{Max}\, u_{dir_{r,r',t}} \qquad \forall r,r' \in J_r, t \end{gathered} \tag{12}$$

$\Delta spu_{r,r'}^{Max}$ is the maximum difference of square pressures for unprocessed gas pipeline segments, usually given by the difference of the square pressure of the produced gas at the wellheads and the square of the minimum pressure at the inlet of a gas processing facility.

Given that $MaxFlowU_{r,r',d,t}$ is a nonnegative variable, we can square eq. (10) leading to the quadratic constraint (10').

$$MaxFlowU_{r,r',d,t}^2 \leq \varphi^2 \;\; diam_d^{5.334} \Delta PU_{r,r',t}^{sq} \qquad \forall r,r' \in J_r, d, t \tag{10'}$$

Finally, maximum gas flowrates in any direction can only be positive if such pipeline of diameter *d* has been installed between *r* and *r'* in previous periods, as imposed by constraint (13). The maximum flow is limited by eq. (14). Constraints (13) and (14) replace eq. (3) for unprocessed shale gas flows.

$$\begin{gathered} MaxflowU_{r',r,d,t} + MaxflowU_{r,r',d,t} \leq \varphi \;\; diam_d^{2.667} \left(\Delta spu_{r,r'}^{Max}\right)^{0.5} \sum_{\substack{t' \in TI \\ t' \leq t}} u_{k,r,r',d,t'} \\ \forall k = gas, r, r' \in J_r, r < r', t \end{gathered} \tag{13}$$

$$Q_{k,r,r',t} \leq \sum_d MaxflowU_{r,r',d,t} \qquad \forall k = gas, r, r' \in J_r, t \tag{14}$$

Analogous constraints are developed to size the pipelines and managing the processed gas flows leaving the processing units to reach the delivery nodes.

$$v_{dir_{r,r',t}} + v_{dir_{r',r,t}} = 1 \qquad \forall r, r' \in J_r, t \tag{15}$$

$$\Delta PP^{sq}_{r,r',t} \le (PP^{sq}_{r,t} - PP^{sq}_{r',t}) + \Delta spu^{Max}_{r,r'} (1 - v_{dir_{r,r',t}})$$
$$\Delta PP^{sq}_{r,r',t} \le \Delta spp^{Max}_{r,r'} \, v_{dir_{r,r',t}} \qquad \forall r, r' \in J_r, t \tag{16}$$

$$MaxFlowP^2_{r,r',d,t} \le \varphi^2 \; diam_d^{5.334} \Delta PP^{sq}_{r,r',t} \qquad \forall r, r' \in J_r, d, t \tag{17}$$

$$MaxflowP_{r',r,d,t} + MaxflowP_{r,r',d,t} \le \varphi \; diam_d^{2.667} \left(\Delta spp^{Max}_{r,r'}\right)^{0.5} \sum_{\substack{t' \in TI \\ t' \le t}} v_{k,r,r',d,t'} \tag{18}$$
$$\forall k = gas, r, r' \in J_r, r < r', t$$

$$R_{k,r,r',t} \le \sum_{d} MaxflowP_{r,r',d,t} \qquad \forall k = gas, r, r' \in J_r, t \tag{19}$$

Equations (18) and (19) replace eq. (6) for processed shale gas flows. Summarizing, the MIQCP for the optimal design of the gas gathering network aims to minimize function (8), subject to constraints (1), (2), (4), (5), (7), (10'), and (11) to (19). Note that the only nonlinear (quadratic) constraints are (10') and (17), which are convex, thus yielding an MIQCP with a convex relaxation.

### *MIQCP Model for the Integrated Design of Shale Oil and Gas Gathering Networks*

It is interesting to note that the two models in the previous sections are totally separable and may be solved in parallel to yield the optimal networks for shale oil and shale gas, respectively. Nevertheless, accounting for production curtailment, a typical operation in unconventional production, leads to the need of integrating both models. Operators curtail production to reduce oil and gas outflows from a well, usually due to a temporary shortage of transportation, processing or delivery capacity. However, under a rather conservative assumption, the curtailed production might be directly lost. We introduce the variable $f_{w,r,t}$ representing the fraction of the forecasted production actually gathered from the well *w* during period *t*. When limiting the production of a well, oil and gas flows are commonly curtailed in the same proportion, and that is why variable $f_{w,r,t}$ lacks the index *k* and is the same for both components. The new eq. (20) replaces eq. (1) to let the model simultaneously decide on the optimal design for shale oil and shale gas networks, and the most convenient plan of curtailment operations.

$$\sum_{w \in W_r} f_{w,r,t} \, p_{k,w,r,t} + \sum_{r' \in J_r} Q_{k,r',r,t} = P_{k,r,t} + \sum_{r' \in J_r} Q_{k,r,r',t} \qquad \forall k, r, t$$
$$0 \le f_{w,r,t} \le 1 \quad \forall r, w \in W_r, t \tag{20}$$

Besides, the optimal sizing and installation planning of two-phase separators within each row (see Figure 2) can be also added to the integrated formulation, as in eq. (21). Similar to eq. (2), the parameter $sc_{k,s}$ accounts for the capacity of a two-phase separator of size *s* to process component *k*, while $z_{k,s,r,t}$ is a 0-1 variable taking value 1 if a two-phase separator of capacity *s* is installed in row *r* at period *t*.

$$\sum_{w \in W_r} f_{w,r,t} \, p_{k,w,r,t} \le \sum_{\substack{t' \in TI \\ t' \le t}} \sum_{s} sc_{k,s} \, z_{k,s,r,t'} \qquad \forall k, r, t \tag{21}$$

Finally, a new objective function is proposed in eq. (22) to account for the actual incomes from oil and gas deliveries, according to the curtailment plan. Eq. (22) also deducts the investment and operating costs of two-phase separators, and the *NPC* of shale oil and shale gas gathering networks, obtained from eq. (8).

$$Max\ NPV = \sum_t \frac{1}{(1+i)^{t-1}} \left[ \sum_{k,r} price_{k,t}\ D_{k,r,t} - \sum_{k,s,r} isf_{k,s}\ z_{k,s,r,t} - \sum_{k,r} osf_{k,r} \sum_{w \in W_r} f_{w,r,t}\, p_{k,w,r,t} \right] - NPC_{oil} - NPC_{gas} \tag{22}$$

Therefore, the MIQCP model for the integrated design of shale oil and gas gathering networks accounting for production curtailment seeks to maximize eq. (22), subject to constraints (2) to (7), (10'), and (11) to (21).

### *Stochastic Programming Models for Combined Shale Oil and Shale Gas Development Strategies*

The deterministic approaches presented in previous sections can be extended to account for uncertain price forecast scenarios and, based on them, alternative well development plans. By adding the index $\omega$ (scenarios) to all model decision variables, one may derive multistage stochastic programming (MSSP) counterparts of the MILP and MIQCP formulations (Li and Grossmann, 2021). In the most general case, each time period $t$ represents each of the stages of the decision making process. Once the uncertainty in oil and gas prices for period $t$ realizes, the development plan is revised by the company accordingly. Based on these facts, the facility planner seeks to make optimal recourse actions for period $t$+1. Once again, after the uncertainty realization at period $t$+1 the facility planner determines recourse actions for $t$+2, and so on and so forth until reaching the end of the time horizon. It is widely known that the size of the MSSP formulations increases exponentially with the number of scenarios and stages, becoming intractable even for relatively small instances.

In this work, we follow an alternative approach to approximately solve the MSSP, which has been originally proposed by Balasubramanian and Grossmann (2004). More specifically, we propose to solve the deterministic equivalents of a series Two-Stage Stochastic Programming (TSSP) models within a shrinking-horizon framework. In the first step (first TSSP model), the first stage, here-and-now decisions are those investments to be made in the first time period, for which we impose the so-called non-anticipativity constraints (NAC), as in eq. (23).

$$\begin{aligned}
x^{\omega}_{k,s,r,t1} &= x_{k,s,r,t1} && \forall k,s,r,\omega \\
y^{\omega}_{k,e,r,t1} &= y_{k,e,r,t1} && \forall k,e,r,\omega \\
z^{\omega}_{k,s,r,t1} &= z_{k,s,r,t1} && \forall k,s,r,\omega \\
u^{\omega}_{k,r,r',d,t1} &= u_{k,r,r',d,t1} && \forall k,r,r' \in J_r, r < r', d, \omega \\
v^{\omega}_{k,r,r',d,t1} &= v_{k,r,r',d,t1} && \forall k,r,r' \in J_r, r < r', d, \omega
\end{aligned} \tag{23}$$

All other decisions (for later periods *t* = 2…*T*) are recourse actions to be optimally determined according to the development plan and revisions. The objective function of the deterministic equivalent formulation can be stated as in eq. (24), where $\tau_\omega$ is the probability of each scenario $\omega$.

$$Max\ ENPV = \sum_{\omega} \tau_\omega\ NPV_\omega \tag{24}$$

Note that the main parameters changing with the scenario $\omega$ are the oil and gas price forecasts ($price_{k,t}^{\omega}$, in eq. 22)) and the expected production from the wells over the time horizon ($p_{k,w,r,t}^{\omega}$, in eqs. 1, 20, 21 and 22) derived from the development plan associated with scenario $\omega$. In a hypothetic scenario in which oil prices are more attractive than gas prices, wellpads with lower GOR will be developed earlier, showing larger values of $p_{k,w,r,t}^{\omega}$ for earlier time periods $t$. The model is iteratively solved for the time horizons *t* = 2…*T*, *t* = 3…*T*, … , *t* = *T*-1…*T*; determining in each step the optimal recourse decisions to be made under every possible scenario observed at period $t$.

## ***Results and Discussion***

In this section we address several case studies where we implement the optimization models described in previous sections. Most of them are based on an illustrative example with 8 rows (72 wellpads) to be developed in the next 6 years, which is represented in Figure 4. Finally, a real-world case study from the Vaca Muerta unconventional formation (Argentina) is addressed at the end of the section. In the first two case studies we assume deterministic conditions, that is, oil and gas future prices as well as the company development plan are given data. In Case study 1 the operator is initially interested on shale oil resources. In other words, we assume that the company plans to develop the oil area first and then move to the gas region. Case 2 presents the opposite situation in which shale gas wellpads are developed first. For each of these illustrative case studies we solve: (a) the MILP model to find the optimal network for gathering shale oil, (b) the MIQCP model to find the optimal network for gathering shale gas, and (c) the extended MIQCP model to simultaneously determine shale oil and shale gas gathering networks, accounting for the possibility of production curtailment. In Case study 3 we address the same problem under uncertain conditions, while Case study 4 presents the real-world problem from the unconventional industry in Argentina. All case studies are implemented on GAMS 37.1 and solved using Gurobi library version 9.5.0, on an Intel Core i7-10510U CPU with 16GB RAM, with 4 parallel threads.

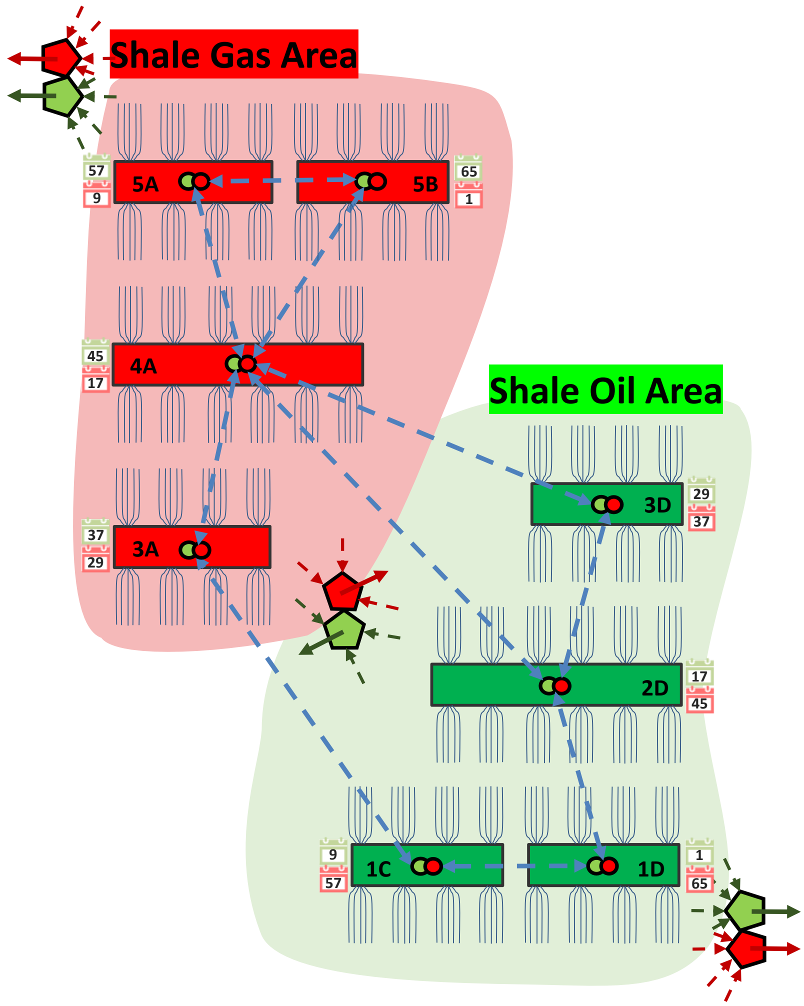


*Figure 4. Illustrative case study comprising a shale gas area (GOR ≈ 20 kscf/bbl) and a shale oil area (GOR ≈ 3 kscf/bbl), with four rows each. Circles illustrate oil and gas collection nodes along each row, while dotted lines are potential pipeline connections. Numbers next to the rows represent starting times for oil and gas production, according to the development plan (green for shale oil focused, red for shale gas focused)*

***Case Study 1: Surface Facility Optimization for a Development Strategy Focused on Shale Oil***

This case study is based on the expected productivity profiles illustrated in Figure 5, according to a development plan focused on shale oil. Oil and gas prices are set at 80 USD per bbl (barrel) and 3 USD per kscf (thousand standard cubic feet) respectively, which stay constant all along the time horizon comprising 6 years (72 months). Monthly production from every wellpad is shown in the Supporting Information. Potential pipeline connections between the rows are shown in Figure 4. There are three alternative pipeline diameters to be used, both for oil and gas flows: 10, 30 and 50 inches. The cost of the pipelines is set at 0.045 MMUSD per inch of diameter and km of length. There are four alternative sizes for primary separation, oil processing and delivery facilities, while gas processing and delivery facilities have five alternative sizes each. Information about the capacity and cost of these facilities is also given in the Supporting Information. Operating costs are assumed to be independent of the network design.

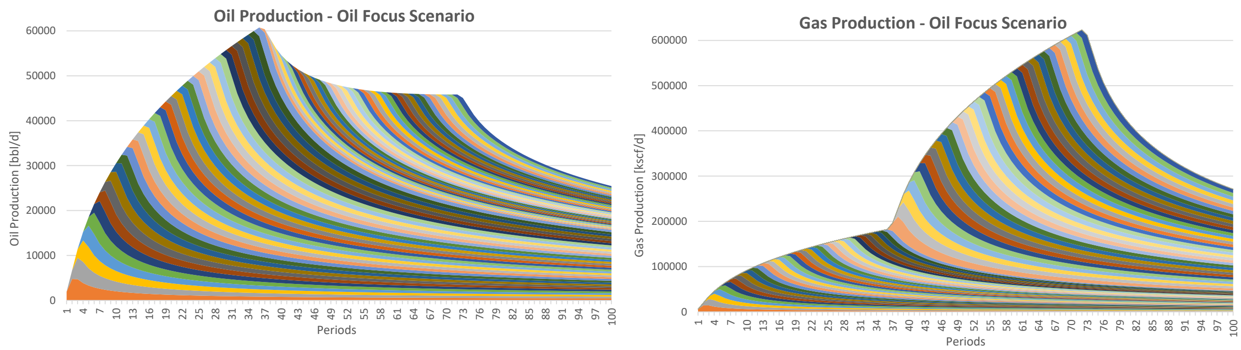


*Figure 5. Expected oil and gas production profiles along the time horizon for a development plan aimed to produce shale oil first. Every color represents the production coming from a different wellpad.*

*Shale Oil Gathering Network*

The size of the MILP model for the shale oil network is reported in Table 1. It is solved to optimality (0% gap) in less than 1 s of CPU, yielding a minimum net present cost of 183.21 million USD including installation of pipelines, processing and delivery facilities. The optimal network just comprises pipelines of 10”, and every facility is located in the shale oil area (row 2D). Both processing and delivery facilities are centralized at that node, noting that oil delivery capacity is expanded in the third year (month 25).

*Shale Gas Gathering Network*

The MIQCP model for the shale gas network is solved to optimality (0% gap) in 440 s of CPU, yielding a minimum net present cost of 941.20 million USD. The optimal network mainly comprises pipelines of 30”, with the exception of branch 1C-1D, with a diameter of 10”. Note that every gas processing and delivery facility is located in the shale gas area (row 5D), with a delivery capacity expansion in the fourth year (month 37).

*Integrated Solution with Production Curtailment*

We finally solve the integrated MIQCP model in which we optimally determine oil and gas gathering networks together with the production curtailment strategy. Due to its higher complexity, the computational time increases to 4717 CPU s. It is interesting to note that, in contrast to the previous case, the model suggests decentralizing the gas processing facilities in two locations: 1D (shale oil region) and 4A (shale gas region), as shown in Figure 7. Moreover, the size of the facilities and their expansions are smaller, given that some of the production coming from the wellpads is conveniently curtailed. This is made to avoid heavy capital investments that would have been used only for a short period of time. Indeed, by resigning 11.85 MMUSD in oil and gas production, investment costs can be reduced by 69.35 MMUSD, with major savings in oil processing and delivery facilities. A comparison of the oil processing and delivery capacities in contrast to the overall production rates over time is shown in Figure 8.

*Table 1. Model statistics and computational results for the illustrative case studies*

| | MILP/MIQCP Separate Models | | MIQCP Integrated Model w/Curtailment | | Two Stage Stochastic Formulation (First Step) |
|---|---|---|---|---|---|
| | Oil Focus Scenario | Gas Focus Scenario | Oil Focus Scenario | Gas Focus Scenario | Scenario 1 Oil Focus - Scenario 2 Gas Focus |
| NPV [MM USD] | 4715.3 | 4879.5 | 4772.9 (+57.5) | 4909.7 (+30.2) | 4781.2 |
| CPU time [s] | 1+440 | 2+33 | 4717 | 1052 | 23697 |
| Equations* | 1783/13933 | 1760/13422 | 16443 | 16758 | 44311 |
| Quadratic Constraints* | 0/3726 | 0/3600 | 3888 | 3888 | 11664 |
| Positive Variables* | 1206/6786 | 1160/6488 | 8503 | 8503 | 22560 |
| Binary Variables* | 336/1650 | 336/1513 | 2275 | 2306 | 6025 |
| Total GAS processing capacity | 625 | 625 | 608 | 480 | 625 |
| NPC of gas proc. facilities [MM USD] | 435 | 435 | 429.9 | 401.7 | 435 |
| Total GAS delivery capacity | 640 | 625 | 640 | 480 | 625 |
| NPC of gas deliv. facilities [MM USD] | 449.3 | 450 | 449.3 | 350 | 450 |
| Total OIL processing capacity | 80 | 80 | 60 | 80 | 80 |
| NPC of oil proc. facilities [MM USD] | 80 | 80.9 | 68.2 | 80.9 | 80 |
| Total OIL delivery capacity | 90 | 90 | 60 | 90 | 90 |
| NPC of oil deliv. facilities [MM USD] | 85.8 | 76.1 | 60 | 74.5 | 85.8 - 82.1 |
| NPC of pipelines [MM USD] | 74.3 | 68.7 | 67.9 | 68.7 | 81.6 - 82.1 |
| Total Production Losses [MM USD] | 0 | 0 | 11.85 | 134 | 6.6 - 11.5 |
| Gas Incomes NPV [MM USD] | 1972 | 4860.3 | 1967.3 | 4767.1 | 1955.8 - 4831.4 |
| Oil Incomes NPV [MM USD] | 7326.6 | 4608.2 | 7319.4 | 4567.2 | 7323.1 - 4604.7 |

**After preprocessing*

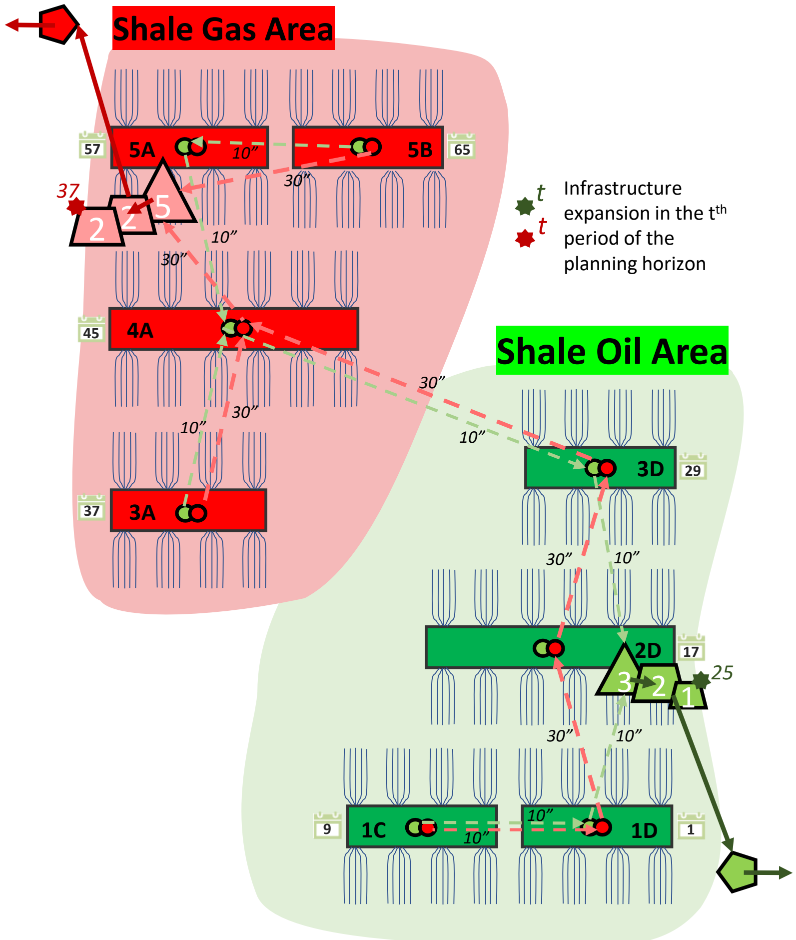


*Figure 6. Optimal solutions found by the MILP and MIQCP models for shale oil and shale gas gathering networks for Case 1, respectively. Development strategy is focused on shale oil first.*

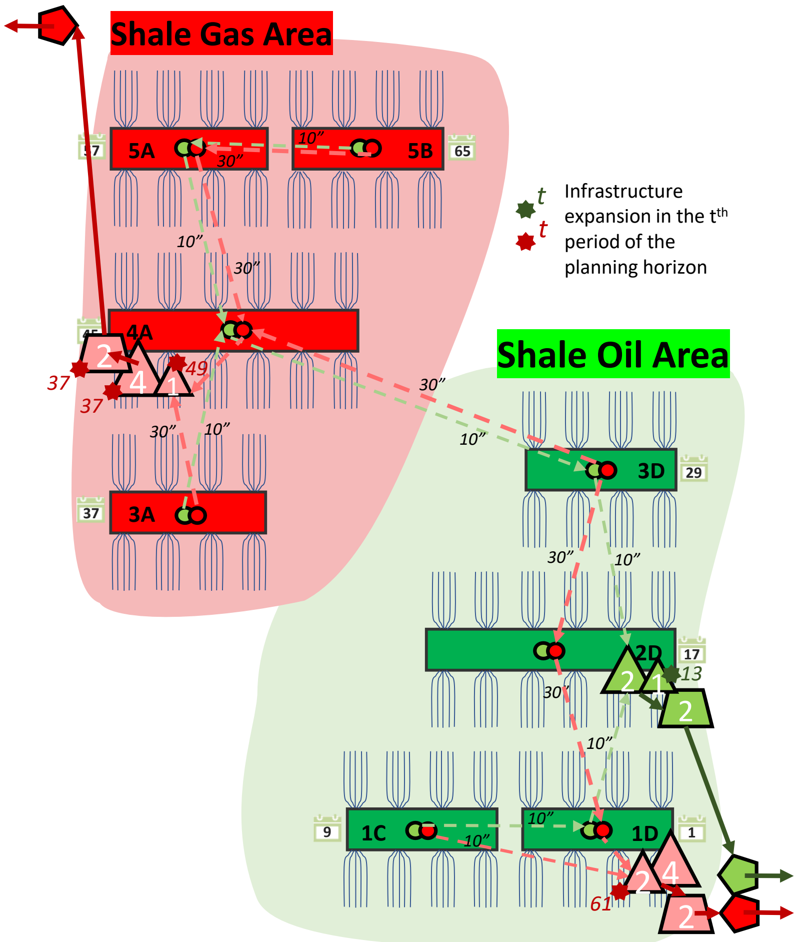


*Figure 7. Optimal solution for Case study 1 found by the integrated MIQCP model, accounting for production curtailment. NPV of the project increases by 57.5 million USD.*

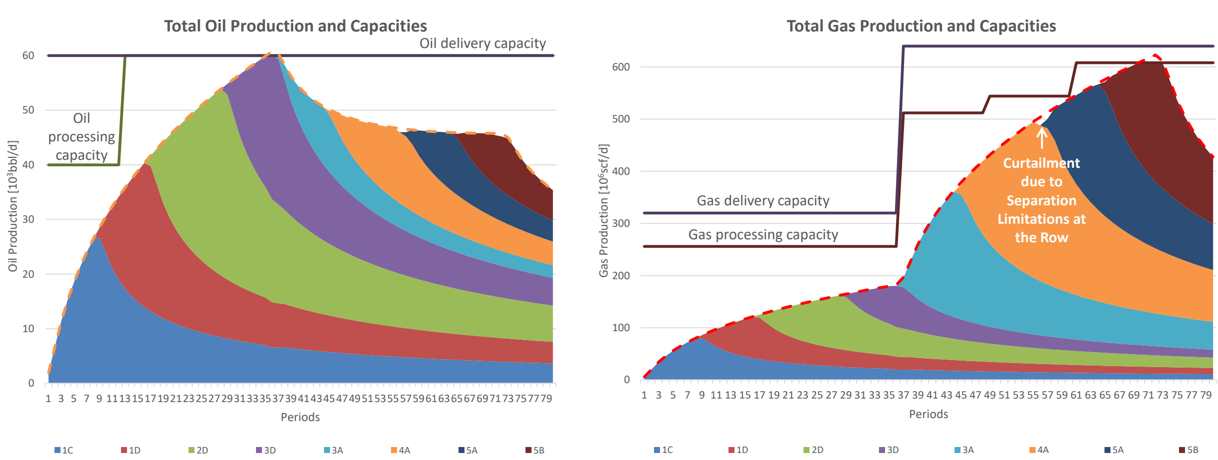


*Figure 8. Total production, processing and delivery capacities in the optimal solution for Case study 1 found by the integrated MIQCP model, accounting for production curtailment.*

*__Case Study 2: Surface Facility Optimization for a Development Strategy Focused on Shale Gas__*

In contrast to the previous case, in Case Study 2 we assume that the operator is interested in developing shale gas resources first (see red numbers next to the rows in Figure 4, representing start times in months), yielding the production profiles shown in Figure 9. This strategy is based on a different scenario for future oil and gas prices: 60 USD per barrel of oil and 6 USD per million BTU of natural gas, for the next 10 years.

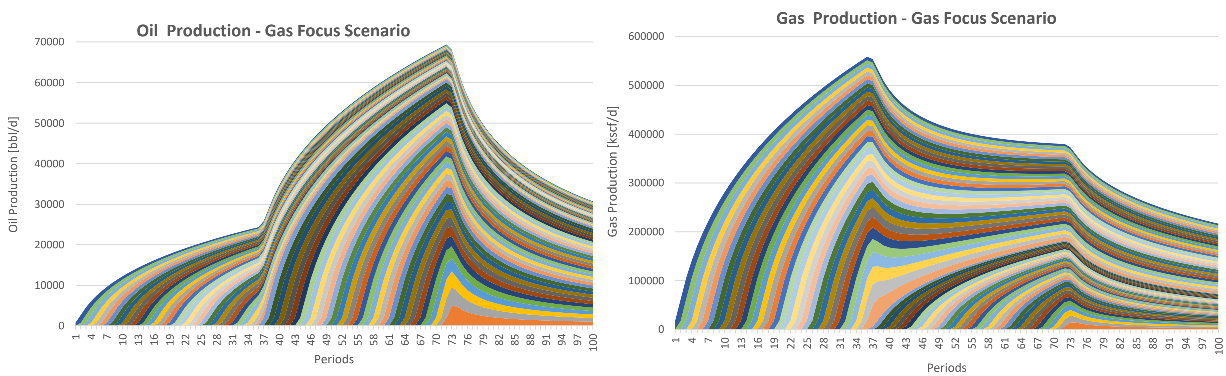


*Figure 9. Expected oil and gas production profiles along the time horizon for a development plan aimed to produce shale gas first. Every color represents the production coming from a different wellpad.*

*Shale Oil Gathering Network*

Computational details of the MILP and results for this instance are given in the Supporting Information. The shale oil network has decentralized processing and delivery facilities at row 5A (shale gas area, built at the initial time) and 2D (shale oil area, built in the 4$^{th}$ year). The net present cost of pipelines, processing and delivery facilities for shale oil is 174.17 MMUSD.

*Shale Gas Gathering Network*

Contrarily to the shale oil network, gas processing and delivery facilities are centralized at row 4A (shale gas region), with large capacities installed at the initial time and showing no further expansions along the time horizon. The NPC of the shale gas network amounts to 936.5 MMUSD. The resulting network can be found in the Supporting Information.

*Integrated Solution with Production Curtailment*

Benefits from production curtailment are smaller in comparison to the previous case. The optimal solution to the integrated problem (MIQCP) is found in 1052 CPUs, yielding an NPV of 4909.7 MMUSD. In turn, the NPV yielded by the separate models for shale oil and shale gas networks with no curtailment is 30.2 MMUSD below that value. Although the shale oil network is the same as the one found by the separate MILP, significant reductions in the size of the shale gas facilities are proposed by the integrated model at the expense of an important production curtailment over months 26 to 41, as shown in Figure 10. More details can be found in the Supporting Information.

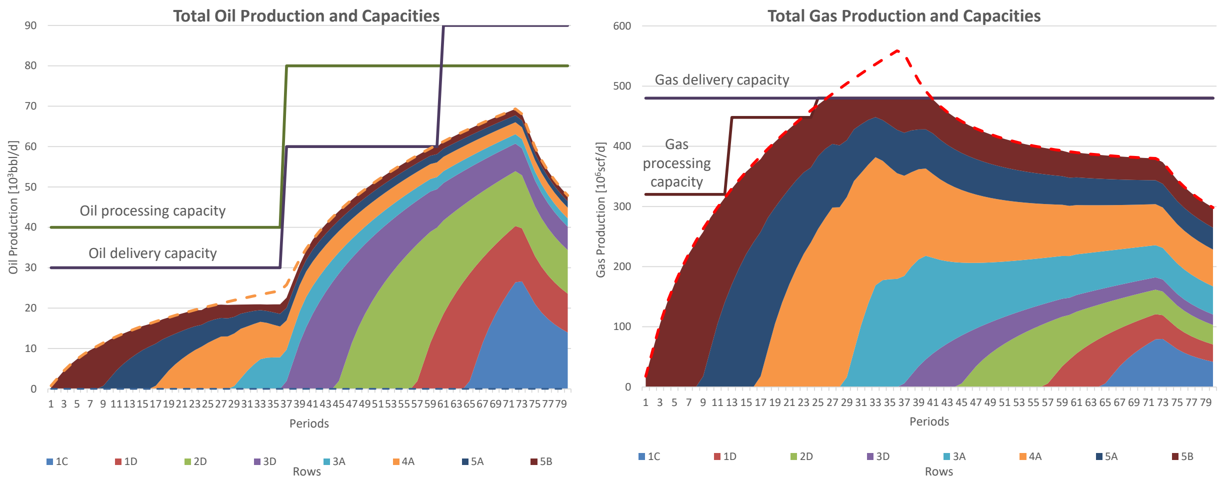


*Figure 10. Total production, processing and delivery capacities in the optimal solution for Case study 2 found by the integrated MIQCP model, accounting for production curtailment.*

***Case Study 3: Surface Facility Optimization under Uncertain Development Strategies***

This case study addresses one of the most important questions faced by facility planners: how to optimally build and expand the gathering networks when the wellpad development strategy is uncertain. From the previous case studies, one can easily infer that imminent decisions (i.e., those of the first year) are clearly different if either the shale oil or the shale gas regions are to be developed first. More specifically, Case Study 1 suggests that if the strategy is focused on shale oil, and production curtailment is allowed, initial facilities of medium size for oil and gas should be placed in the shale oil region. On the contrary, if shale gas is the most valuable resource, facilities for oil and gas should be first built in the shale gas area. But what is the best location, size and expansion plan for processing and delivery facilities if future oil and gas prices (and related strategies) are uncertain? Under simplifying assumptions, this problem can be formulated as a multi-stage stochastic programming model. Let us assume that, at the first day of every year, the company decides to develop wellpads from the shale oil or the shale gas region. The order and pace to develop the wellpads in each of both regions is known, but the operator may mobilize rigs and fracturing crews from one region to the other driven by more attractive prices of oil or gas (chasing price).

Figure 11 illustrates the scenario tree related to the problem presented in this example. At the initial time, the facility planner is not sure if the company will finally start the development plan over the oil (green arrows in Figure 11) or the gas region (red arrows). For simplicity, let us assume that probabilities are 50% for any of them. Given that facilities installations need to be made in advance, the gathering network might be ready to cope with any of both scenarios. After the first year, the operator faces a new decision: stay (usually, more likely) or mobilize. Once again, the facility planner needs to make recourse actions even before the operator decides on one or the other course. And the process is repeated every year. Following the stochastic programming approach presented in previous sections, we address this problem by sequentially solving the deterministic equivalents of two-stage stochastic programming formulations, over shrinking horizons.

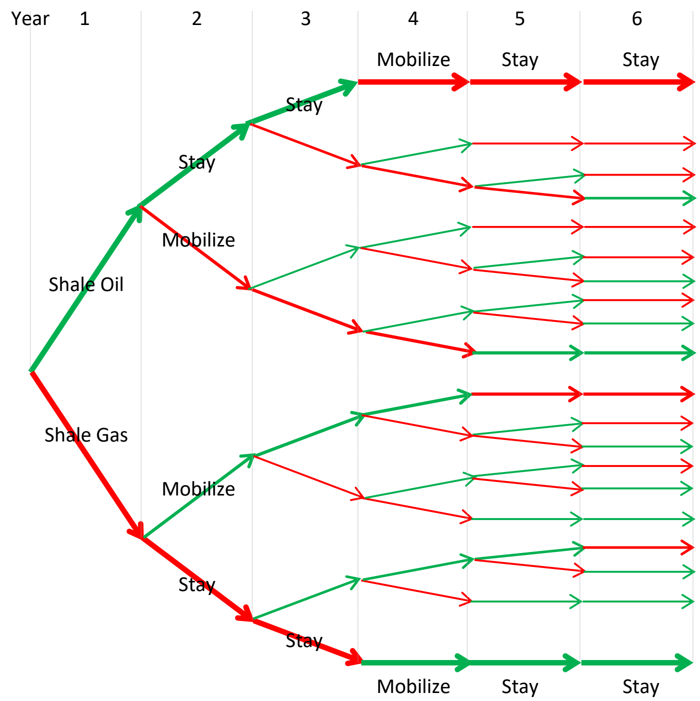


*Figure 11. Scenario tree for Case Study 3, assuming an uncertain development plan. Key question: stay or mobilize to shale oil / shale gas region.*

*Step 1*

At the first step, investments of year 1 are first-stage, here-and-now decisions. All other investments (for years 2, 3, 4, 5 and 6) are second-stage, wait-and-see decisions. The integrated MIQCP model for step 1 should account for 20 scenarios (leaves of the scenario tree depicted in Figure 11). However, the resulting model is computationally intractable. From that, we suggest to reduce the set of scenarios to the two most likely and representative cases, i.e., the upper and lower branches in Figure 11. More specifically: (a) start at the shale oil area, finish and mobilize to shale gas area; or (b) start at the shale gas area, finish an mobilize to the shale oil area. Although this scenario tree pruning may hinder the solution of the multi-stage stochastic program, finding the global optimum is not the aim of this case study. Our goal is to obtain indicators to conclude on the actual value of stochastic solutions, and how they impact the planning of the gathering networks.

The optimal solution to this step is found after 6.5 hours of computation (see Table 1). The MIQCP model comprises more than 44,000 constraints, 22,000 continuous variables and 6,000 discrete variables, even after solver preprocessing. The optimal solution suggests a different network in comparison to the deterministic cases when production curtailment is allowed: initial facilities for oil are placed in the shale oil region (row 2D), while initial facilities for gas are installed in the shale gas region (row 4A). Moreover, the initial size of these facilities is large enough to process oil and gas in any of both scenarios during the first two years. From that, there are only two recourse actions over the planning horizon. The oil processing capacity needs to be expanded in the 3$^{rd}$ year, if the strategy is first focused on oil, while the same expansion is required in the 5th year, if the gas scenario realizes. Further details on this solution can be found in Table 1 and in the Supporting Information.

*Step 2*

Having determined the optimal stochastic solution for year 1, the next step is to fix the initial investment decisions and solve two different two-stage stochastic programming models, according to each of the possible outcomes of the first year. In other words, the next step is to determine the optimal recourse actions to be made at the start of the second year by knowing which of the areas have been first developed over the first year. We assume that the probability of staying at the current area is 75%, while mobilizing stands for the other 25%. Interestingly, there is no need to make recourse actions at this time, under any of both scenarios. The need to expand the oil processing capacity is first observed at the third iteration, only if the operator decides to start and stay at the shale oil region. Optimal solutions to the deterministic equivalents for years 2 to 6 are found in less than 3 seconds of CPU.

*Stochastic Indicators*

According to Birge and Loveaux (2011) the value of the stochastic solution (VSS) is the difference between the recourse problem solution (expected NPV from the stochastic solution) and the expected value solution (EEV). In our problem, the EEV can be thought of as the NPV yielded by the optimal gathering network to be designed if the strategy followed by the company is mixed, i.e., oil and gas prices are at average levels, and both regions are developed evenly and simultaneously, using the available resources. More specifically, we find the EEV by solving the integrated MIQCP model allowing curtailment, based on an average development plan in which the oil and gas prices are 70 USD per barrel and 4.50 USD per MMbtu, respectively. Moreover, the pace of activity in each region (number of wells developed per year) is reduced by half in comparison to Case Studies 1 and 2, following the same order in both regions. We finally obtain a VSS = 4781.2 - 4766.7 = 14.5 MMUSD, meaning that the company may increase the benefits by 14.5 MMUSD (expected value) if it follows the investment plan suggested by the stochastic solution instead of relying on the solution yielded by an average scenario.

Also, by comparing the solution to Case Studies 1 and 2 towards the stochastic solution obtained in Case Study 3, we obtain an estimation of the value of perfect information (VPI). By comparing the optimal solutions under perfect information (i.e., assuming that we know that the focus will be first on oil or gas) with the best solutions that will be obtained under both scenarios if the starting network is built as explained in Case Study 3, we obtain VPI = 0.5* (4772.9 - 4708.1) + 0.5*(4909.7 - 4854.4) = 60.0 MMUSD. In future work we will address the impact of uncertainty with more detail.

***Case Study 4: Real-World Example from the Unconventional O&G Industry***

A real-world problem from the Vaca Muerta shale formation of Argentina is finally addressed in this section. For confidentiality reasons, we do not provide details on the actual development plan from the company. However, we illustrate the geographical distribution of the rows of wellpads and the order in which they should be developed under the current scenario of oil and gas prices (see Figure 12). There are three sectors in the map: shale oil (to be developed first), transition and shale gas sectors. The numbers next to each row account for the order of development over a planning horizon comprising 450 periods. The plan is targeted to an overall production of more than 100,000 barrels per day of oil, and as rigs and frac crews move to transition and shale gas regions, natural gas production increases from 150,000 to 250,000 kscf per day. There are 4 alternative locations to deliver processed shale oil and shale gas flows, and any of the 15 rows is a candidate location to install (and expand) processing and/or delivery facilities for both fluids.

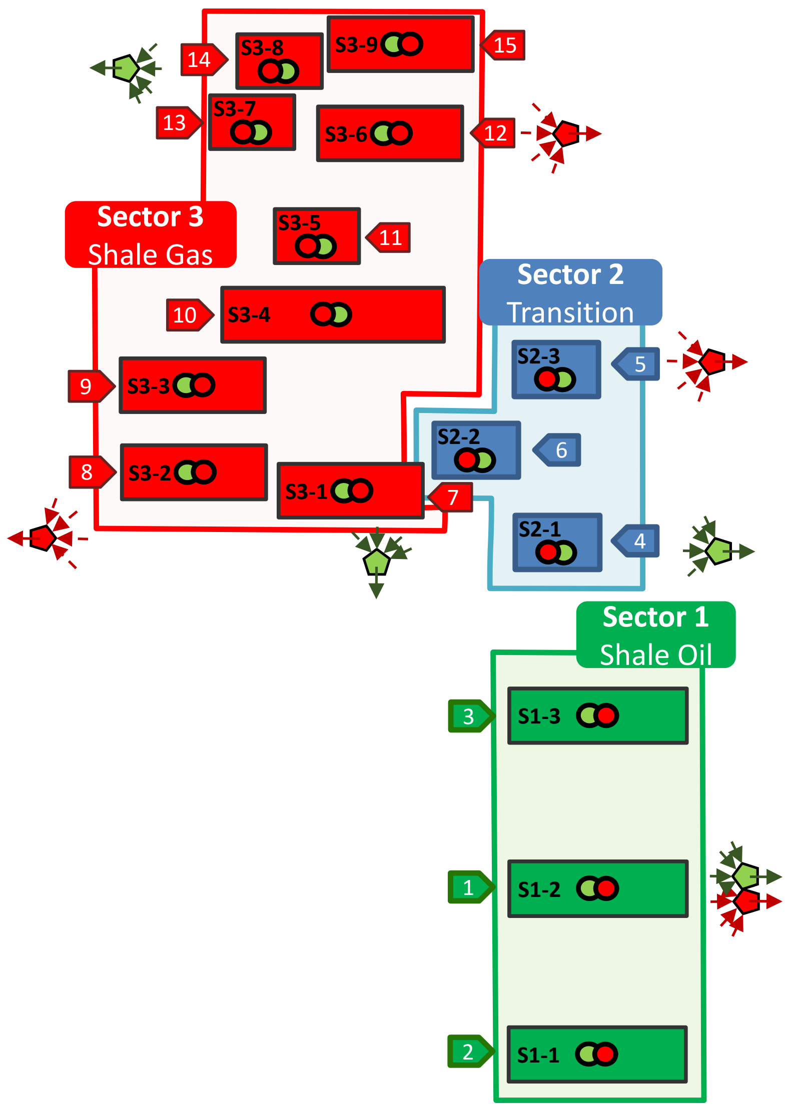


*Figure 12. Illustration of the rows of wellpads to develop in the real-world example (Case Study 4)*

Given that curtailment is allowed in this case, we solve the integrated MIQCP model to simultaneously determine the optimal gathering networks for unconventional oil and gas. After preprocessing, the model comprises more than 70,000 constraints (of which 17,712 are quadratic) and 48,000 variables (of which 13,677 are integer). After 40,000 seconds of computation, the global optimality gap is reduced to

0.17%. The best solution found is illustrated in Figure 13. As depicted, the optimal decision is to initially build oil and gas facilities in a centralized location (row S1-2, i.e. the first row to be developed). However, different strategies are followed to expand oil and gas facilities. Gas processing facilities are expanded five times along the time horizon (see Figure 14), while gas delivery facilities require a single expansion at time 289. All of them in the same centralized location. In contrast to gas, new oil processing and delivery facilities are installed in a different node (row S3-4) after 120 periods of operation, when drilling and fracturing resources move to the shale gas region. In general terms, the size of the facilities and expansions is rather small, which is justified by non-significant economies of scale and a stringent discount rate for the economic evaluation of the project (10% annually). In Figure 14 we illustrate the evolution of the gas flows and processing capacities over time. Note that no curtailment is required at any time over the planning horizon.

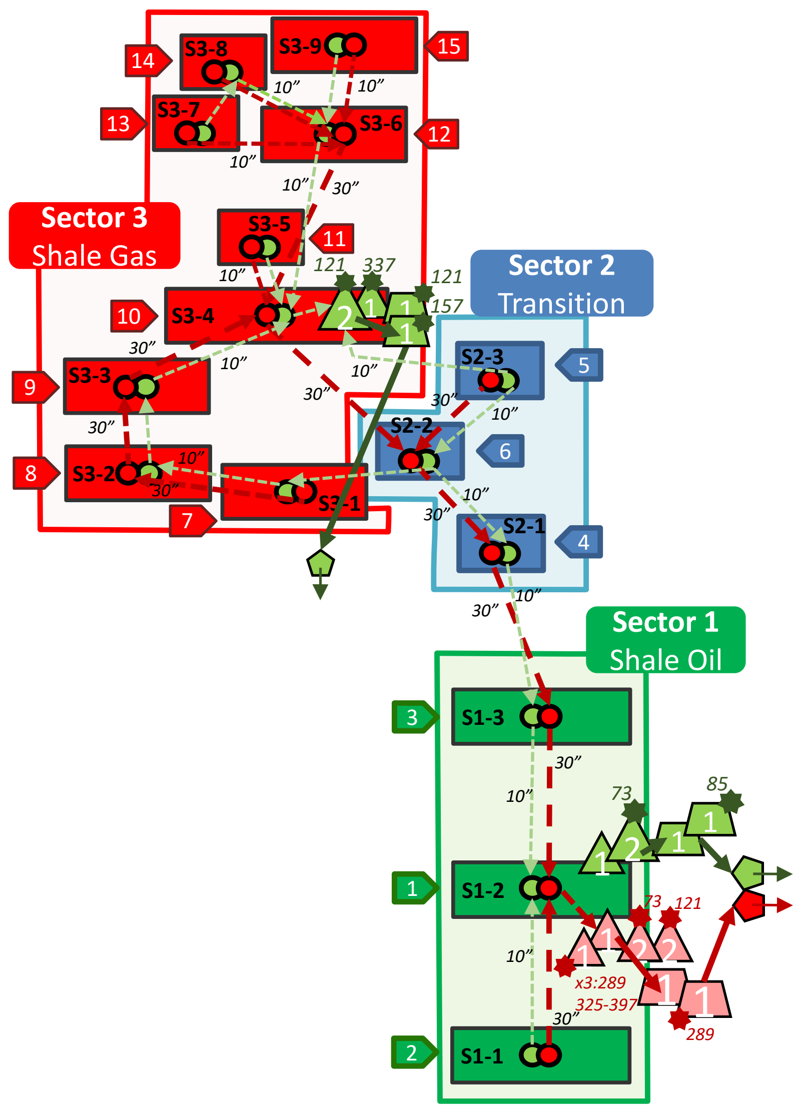


*Figure 13. Best solution found for Case Study 4, illustrating shale oil and gas gathering networks and expansions.*

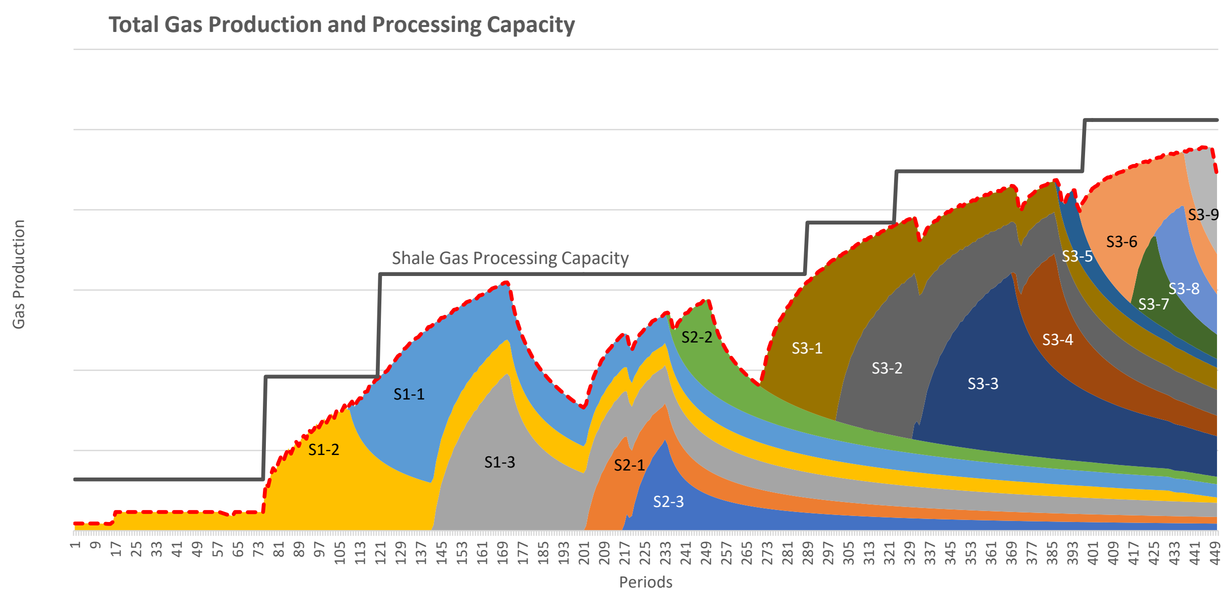


*Figure 14. Expected shale gas production profile coming from the rows and overall gas processing capacity, following optimal expansions.*

## *Conclusions*

We have developed multiperiod, superstructure-based mathematical formulations to address the optimal design of the network of surface facilities gathering, processing and delivering shale oil and gas in combined development strategies. First, we present a mixed-integer linear programming (MILP) model for the shale oil gathering network design; then a mixed-integer, quadratically constrained (MIQCP) model for the shale gas network design; and finally an MIQCP model for the integrated design of gathering networks accounting for production curtailment. A multi-stage stochastic programming model and a solution strategy are afterwards proposed to address combined shale oil and gas development strategies under uncertain scenarios. There are many new challenges faced in this work. Among them, (i) the number of echelons in the gathering networks (i.e., the "steps" or pipeline segments in the path from the wells to the delivery nodes) is not given beforehand; (ii) the flow direction in oil and gas pipelines can be reversed in different time periods; and (iii) the elements of oil and gas network formulations need to be combined into a single optimization approach to cope with changing conditions and uncertain development plans.

Using the MILP and MIQCP models we separately find the optimal designs of the oil and gas networks under two alternative scenarios: shale oil-first (Case Study 1) or shale gas-first (Case Study 2). Results demonstrate that under the oil-first scenario, oil and gas processing and delivery facilities should be placed and expanded in centralized locations, each of them in their corresponding area (oil facilities in the oil sector, gas facilities in the gas sector). Contrarily, if the focus is on shale gas wells, oil facilities should be decentralized, i.e. first placed in the shale gas area and then expanded in the oil area. Interestingly, when both gathering networks are optimized simultaneously allowing production curtailment, the net present value of the project increases between 30 and 60 MMUSD by making a better use of the separation and processing capacities. Case Study 3 proves that there may be significant value in the solution yielded by stochastic formulations. The expected NPV from the stochastic solution increases by 14.5 MMUSD when compared to the solution found by a deterministic model based on an average scenario. Moreover, by comparing the solution to Case Studies 1 and 2 towards the stochastic solution obtained in Case Study 3, the expected value of perfect information (EVPI) is 60 MMUSD.

The models have been finally validated by solving a real-world problem from the Vaca Muerta shale formation (Case Study 4). The plan is targeted to an overall production of more than 100,000 barrels per day of oil coming from 15 rows of wellpads, and as rigs and frac crews move to shale gas regions, natural gas production increases from 150,000 to 250,000 kscf per day. We solve the integrated MIQCP model to simultaneously determine the optimal gathering networks for oil and gas, which comprises more than 70,000 constraints (being 17,712 quadratic) and 48,000 variables (being 13,677 integer). After more than 10 hours of computation, the global optimality gap is reduced to 0.17%. The optimal solution suggests to initially build oil and gas facilities in a centralized location, but different strategies are followed to expand oil and gas facilities. Gas processing facilities are expanded five times in the same centralized location. Instead, new oil processing and delivery facilities are installed in a different node after some periods of operation, when drilling and fracturing resources move to the shale gas region. In general terms, the size of the facilities and expansions is rather small, which is justified by non-significant economies of scale and a stringent discount rate for the economic evaluation of the project.

Future work aims to assess the impacts of uncertainty with more detail. New stochastic programming formulations and alternative solution strategies are required to conduct computational experiments accounting for a larger set of possible scenarios with reasonable computational times so as to assess the sensitivity of the optimal facility investment plan. The final goal will be to provide guidelines to efficiently adapt to a changing policy and investment landscape, driven by the energy transition.

## *Acknowledgments*

Financial support from National University of Litoral and CONICET is gratefully acknowledged. Financial support from ExxonMobil Upstream Research Company through the Center of Advanced Process Decision-making at Carnegie Mellon University is also gratefully acknowledged.

## *Nomenclature*

### *Sets*

*D* Set of alternative pipeline diameters

*T* Set of time periods (months)

### *Parameters*

### *Nonnegative Variables*

***Binary variables***

## *References*


1. *Cafaro DC, Grossmann IE (2020). Optimal design of water pipeline networks for the development of shale gas resources. AIChE Journal 67(1). DOI: 10.1002/aic.17058*
2. *Weymouth TR (1912). Problems in Natural Gas Engineering. Transactions of the ASME 34:185-234.*
3. *Li C, Grossmann IE (2021). A Review of Stochastic Programming Methods for Optimization of Process Systems Under Uncertainty. Front. Chem. Eng. https://doi.org/10.3389/fceng.2020.622241*
4. *International Energy Agency. World Energy Outlook 2019, OECD Publishing: 2019.*
5. *US Energy Information Agency. Annual Energy Outlook 2020.*
6. *Weissman S, Constantine S, Hernandez P, Gallagher C (2016) Natural Gas as a Bridge Fuel – Measuring the Bridge, Center for Sustainable Energy, San Diego, CA. March 2016.*
7. *Nunes C, Can Natural Gas Be a Bridge to Clean Energy? National Geographic, Jan 10, 2020.*
8. *Yang L, Grossmann IE, Manno J (2014). Optimization models for shale gas water management. AIChE J., 60: 3490-3501.*
9. *US Environmental Protection Agency. Engineering and Analysis Division. Office of Water. Detailed Study of the Centralized Waste Treatment Point Source Category for Facilities Managing Oil and Gas Extraction Wastes. EPA-821-R-18-004. May 2018.*
10. *Rassenfoss S (2011). From Flowback to Fracturing: Water Recycling Growsin the Marcellus Shale. Journal of Petroleum Technology, 63: 48–51.*
11. *Blauch ME (2010). SPE Unconventional Gas Conference, 23-25 February, Pittsburgh, Pennsylvania, USA. doi:10.2118/131784-MS*
12. *Drouven M, Grossmann IE (2017) Optimization models for impaired water management in active shale gas development areas. Journal of Petroleum Science and Engineering, 156: 983-995.*
13. *Sun Y, Wang D, Tsang DCW, Wang L, Ok YS, Feng Y (2019). A critical review of risks, characteristics, and treatment strategies for potentially toxic elements in wastewater from shale gas extraction. Environment International, 125:452-469.*
14. *US Energy Information Agency. Growth in Argentina's Vaca Muerta shale and tight gas production leads to LNG exports. Today in Energy, July 12, 2019.*
15. *Awalt J (2019). North America Pipeline Construction Outlook. Underground Construction, 74(1), January 2019.*
16. *Congressional Research Service (2014). US Rail Transportation of Crude Oil: Background and Issues for Congress.*
17. *Yang L, Grossmann IE, Mauter MS, Dilmore RM (2015). Investment optimization model for freshwater acquisition and wastewater handling in shale gas production. AIChE J., 61: 1770-1782.*
18. *Carrero-Parreño A, Reyes-Labarta JA, Salcedo-Díaz R, Ruiz-Femenia R, Onishi VC, Caballero JA, Grossmann IE (2018). Holistic Planning Model for Sustainable Water Management in the Shale Gas Industry. Ind. Eng. Chem. Res., 57:13131-13143.*
19. *Ondeck A, Drouven M, Blandino N, Grossmann IE (2019). Multi-operational planning of shale gas pad development Comput. Chem. Eng., 126: 83-101.*
20. *Cafaro DC, Grossmann IE (2020). Logic-based formulation for scheduling problems using the concept of campaigns. Submitted for publication.*

21. *Cafaro DC, Grossmann IE (2014). Strategic planning, design, and development of the shale gas supply chain network. AIChE J., 60(6): 2122-2142*
22. *Drouven M, Grossmann IE (2016). Multi-period planning, design, and strategic models for long-term, quality-sensitive shale gas development. AIChE J., 62(7): 2296-2323.*
23. *Cafaro DC, Cerdá J (2014). Rigorous formulation for the scheduling of reversible-flow multiproduct pipelines Comput. Chem. Eng., 61: 59-76.*
24. *Drouven M, Cafaro DC, Grossmann IE (2017). Stochastic programming models for optimal shale well development and refracturing planning under uncertainty AIChE J., 63(11): 4799-4813.*
25. *Caballero JA, Ravagnani MASS (2019). Water distribution networks optimization considering unknown flow directions and pipe diameters. Comput. Chem. Eng., 127: 41-48.*
26. *Bragalli C, D'Ambrosio C, Lee J, Lodi A, Toth P (2012). On the optimal design of water distribution networks: a practical MINLP approach. Optimization and Engineering, 13: 219-246.*
27. *Araya N, Luca FA, Cisternas LA, Gálvez ED (2018). Design of desalinated water distribution networks: complex topography, energy production, and parallel pipelines. Ind. Eng. Chem. Res., 57(30): 9879-9888.*
28. *Kondili E, Pantelides CC, Sargent RWH (1993), A general algorithm for short-term scheduling of batch operations—I. MILP formulation. Comput. Chem. Eng., 17(2): 211-227.*
29. *Mauter M, Palmer VR (2014). Expert elicitation of trends in Marcellus oil and gas wastewater management. Journal of Environmental Engineering, 140(5): B4014004.*