



# An energy system optimization model accounting for the interrelations of multiple stochastic energy prices

Hongtao Ren<sup>1</sup> · Wenji Zhou<sup>2</sup> · Hangzhou Wang<sup>3</sup> · Bo Zhang<sup>4</sup> · Tiejue Ma<sup>1,5</sup>

Accepted: 11 August 2021 / Published online: 30 August 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

The variation of and the interrelation between different energy markets significantly affect the competitiveness of various energy technologies, therefore complicate the decision-making problem for a complex energy system consisting of multiple competing technologies, especially in a long-term time frame. The interrelations between these markets have not been accounted for in the existing energy system modelling efforts, leading to a distortion of understanding of the market impact on the technological choices and operations in the real world. This study investigates the strategic and operational decision-making problem for such an energy system characterized by three competing technologies from crude oil, natural gas, and coal. A stochastic programming model is constructed by incorporating multiple volatile energy prices interrelated with each other. Oil price is modelled by the mean-reverting Ornstein–Uhlenbeck process and serves as the exogenous variable in the ARIMAX models for natural gas and downstream plastic prices. The K-means clustering method is employed to extract a handful of distinctive patterns from a large number of simulated price projections to enhance the computing efficiency without losing retaining critical information and insights from the price co-movement. The model results suggest that the high volatility of the energy market weakens the possibility of selecting the corresponding technology. The oil-based route, for example, gradually loses its market share to the coal approach, attributed to a higher volatile oil market. The proposed method is applicable to other problems of the same kind with high-dimensional stochastic variables.

**Keywords** Energy system modelling · Stochastic programming · Oil market · k-means clustering · Energy price volatility

---

✉ Wenji Zhou  
wenji.zhou@gmail.com

<sup>1</sup> School of Business, East China University of Science and Technology, Meilong Road 130, Shanghai 200237, China

<sup>2</sup> School of Applied Economics, Renmin University of China, Beijing 100872, China

<sup>3</sup> China Petroleum Planning and Engineering Institute (CPPEI), China National Petroleum Corporation, Beijing 100083, China

<sup>4</sup> SINOPEC Beihai Refining and Chemical Co., Ltd., South of No. 4 Road, Beihai 536016, Guangxi, China

<sup>5</sup> International Institute for Applied Systems Analysis, Schlossplatz 1, 2361 Laxenburg, Austria

## 1 Introduction

Energy systems are subject to substantial market uncertainty. Some extreme events, such as the coronavirus crisis, caused a dramatic plunge in the global oil market in 2020. The complexity of the energy supply chain is reflected by the fact that one product could be produced from a variety of technologies with various fuels and types of feedstock. For instance, downstream petrochemical products such as plastics may come from technologies fueled by oil, natural gas, or coal. As a result, fluctuation in one energy market would pass on its impacts to another. This raises the question of how these interrelations between energy markets affect capacity expansion and plant operations? Risks from prices and demands have been assessed in many supply chain decision-making exercises (Awudu & Zhang, 2013; Borodin et al., 2016). Some studies have also analyzed market uncertainty in particular energy system models (ESM) to examine the effects from energy market variation or price volatility (Kim et al., 2019; Lund et al., 2018; McCollum et al., 2016). Risk measuring methods such as incorporating Conditional Value-at-Risk (CVaR) have been tested in the optimization models (Alonso-Ayuso et al., 2018).

Most of the studies consider single price variables as uncertain or stochastic in the models, yet interactions between multiple markets in a supply chain network remain unaccounted for in the existing modelling efforts. Questions like what are the impacts from the interplay between energy markets on the operations of these technologies at a more granular timescale have not been answered, particularly for the energy systems that involve many competing technologies linking a range of commodities traded in separate but interrelated markets. Stochastic programming models prove an effective tool that enables planners to address uncertainties in some key parameters. However, modelling multiple stochastic variables would be computationally demanding, especially for large-scale energy system models (ESMs).

This study aims to address these problems by investigating how the interrelations across multiple volatile energy markets affect the strategic and operational decision-making in ESM and what consequences in carbon emissions would be. To this end, we construct a stochastic programming model for a petrochemical energy system. The targeted supply system network starts with upstream production technologies with different types of feedstock, namely, coal, crude oil, and natural gas, for downstream plastics production. These factors in the decision-making problem are pretty common when energy firms are coping with production planning and scheduling, as energy commodities are becoming more internationally traded nowadays, and mutual effects are increasingly strong (BP, 2020).

The paper is structured as follows. After this introduction, Sect. 2 overviews the literature in addressing uncertainty in ESMs, and the interrelations between energy prices through econometric and statistical approaches. Section 3 describes the problems as well as the supply chain in the scope. Section 4 details model construction, focusing on how multiple stochastic variables are treated in the model. Section 5 summarizes the results and explores policy implications. Section 6 concludes and discusses possible research directions.

## 2 Literature review

ESMs are essential methods used to generate a range of insights and analyses on the supply and demand system of energy (Pfenninger et al., 2014). One of the ESM modeling approaches is conducted in the way of supply chain optimization, which minimizes the total cost of a multi-layer, intertwined energy supply system with a bunch of constraints.

Dealing with price uncertainty has long been a critical issue in energy system modelling. It lies in two crucial aspects, modelling price variation and incorporation of price variation into ESM. In many cases, these two aspects show disparate research domains. Different price movement patterns have been modelled and analyzed by time-series techniques. However, accounting for the variation in ESM is focused on examining the effects of varying price levels on the results of energy system planning. This section summarizes the existent studies in the literature and highlights how this study attempts to advance in this domain to improve the representation of the interplay between different energy markets in an intertwined energy system.

## 2.1 Modelling price uncertainty in energy supply system optimization

Energy supply system optimization is focused on problems such as capacity configuration, investment planning, unit commitment, and electricity dispatch for power systems, among others (Kim & Poor, 2011; Soroudi et al., 2016; Wei et al., 2015). Many energy markets are emerging amid a trend of deregulation worldwide, leading to high-volatility price variation. Modelling the volatile price movements, as well as the associated demand response, has begun critical to ESM.

There have been various approaches to coping with price variation or other uncertain sources in ESM. Table 1 summarizes some examples of incorporating price uncertainty into different ESMs. Although there are many types of ESM, the majority of the selected studies fall into optimization, reflecting a variety of methods coping with price uncertainty. For example, stochastic programming (SP) models the uncertain data with an essential assumption that the true probability distributions need to be known or estimated. It deals with uncertainty in the way of optimizing the expected value over the possible realizations and is often regarded as a scenario-based approach for optimization under uncertainty (Grossmann et al., 2016). The range of price variation can also be approached by fitting time-series models or other methods to generate a set of representative price scenarios, termed as ‘alternative scenarios’ (Lund et al., 2018). This approach provides a straightforward representation of price movement along the modelling horizon; therefore, it could directly illustrate the effects on the outcome variables (Tables 2, 3, 4).

Robust optimization (RO) represents another paradigm that does not require known probability distributions. It instead assumes that the uncertain data resides in the so-called uncertainty set and is deemed more computationally tractable for many classes of uncertainty sets and problem types (Gorissen et al., 2015). Managing assets investment is an active area where RO is frequently applied to accounting for price uncertainty (Kim et al., 2019). Chance constrained programming is regarded as another powerful tool to cope with fluctuating prices in energy system optimization problems (Ning & You, 2019). It aims to optimize an objective while ensuring constraints to be satisfied with a specified probability in an uncertain environment (Uryasev, 2000). A prominent feature of this approach lies in allowing decision-makers to choose their risk levels and therefore examine the improvement in objectives (Ning & You, 2019). It thus has been employed in numerous applications, including process design and operations and supply chain optimization (Ye & You, 2016; You and Grossmann, 2011).

Nevertheless, it generally has a disadvantage of computational inefficiency when the complexity of the problem increases. Other practices include introducing fuzzy numbers into a diffuse model to improve information capture and representativeness, accounting for uncertainty regarding concepts or linguistic nature (Cunico et al., 2017). It is noteworthy

**Table 1** Selected literature of accounting for price variation in ESM

References	ESM type	Uncertainty sources	Uncertainty modelling method
Kim et al. (2019)	Optimization	Solar photovoltaic price	Stochastic programming
Lund et al. (2018)	Simulation	Fuel prices, electricity price	Alternative scenarios
Soroudi et al. (2016)	Optimization	Electricity price	Robust optimization
Wei et al. (2015)	Optimization	Electricity price	Robust optimization
Ye and You (2016)	Simulation + optimization	Inventory	Chance-constrained
McCollum et al. (2016)	Optimization	Fossil fuel prices	Alternative scenarios
Seljom and Tomasgard (2015)	Optimization	Electricity price	Stochastic programming
Babonneau and Haurie (2019)	Optimization	Electricity price	Robust optimization
Cunico et al. (2017)	Optimization	Fossil fuel prices, availability of fossil reserves, demands	Fuzzy number
Rout et al. (2008)	Optimization	Fossil fuel prices	Alternative scenarios
Ren et al. (2019)	Optimization	Carbon price	Stochastic programming
Kim and Poor (2011)	Simulation + optimization	Electricity price	Stochastic programming

**Table 2** Parameters and data for the upstream technologies of the energy system

Parameter	Unit	Oil-based	Coal-based	Gas-based
Capacity	Mton/yr	1.095	1.095	1.095
Life span	years	25	25	25
<i>Input</i>				
Naphtha	Mton/yr	2.901	–	–
Coal	Mton/yr	–	7.576	–
Ethane-propane	Mton/yr	–	–	1.553
<i>Output</i>				
Ethylene	Mton/yr	0.986	0.986	0.986
Propylene	Mton/yr	0.404	0.862	0.224
Capital investment	Million yuan	4877	35,795	4268
Fixed O&M	Million yuan	147	716	141
Variable O&M	Million yuan	628	2219	526
Economic share of ethylene		0.41	0.53	0.66
CO <sub>2</sub> emissions	Ton/ton	1.13	5.79	1.58

that selecting the most appropriate method depends on the nature of the targeted problem and the uncertainty sources in the scope.

**Table 3** Key data and assumptions for the downstream supply chain

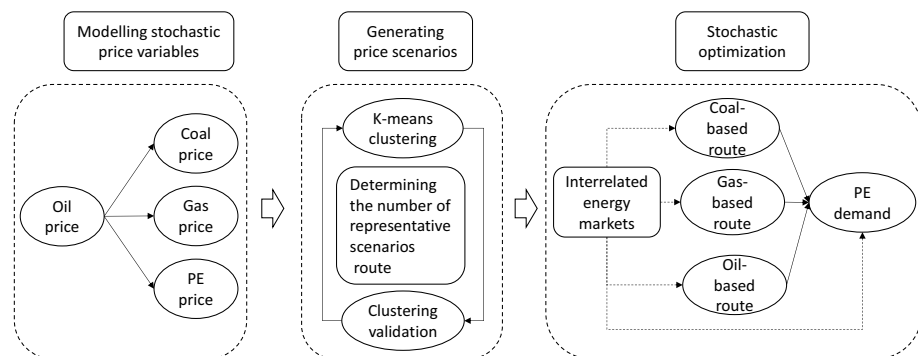
	Unit	Plastic plant 1	Plastic plant 2		
Capacity for Product A	Thousand tons Kton/yr	630	350		
Capacity for Product B	Kton/yr	820	750		
Capital cost	Million yuan	2375.82	2039.47		
Fixed O&M	Million yuan/yr	16.63	14.28		
Variable O&M	Million yuan/yr	2.85	2.45		
	Unit	Customer 1	Customer 2	Customer 3	Customer 4
Demand for product A	Kton/yr	72,163	48,503	83,993	70,980
Demand for product B	Kton/yr	83,993	48,503	106,470	141,960

**Table 4** Validation criteria for different numbers of clusters

	Number of clusters	CP	SP	Ratio of SP and CP
	4	5631	66,088	11.74
	5	5662	70,366	12.43
	6	4738	60,475	12.76
	7	6316	79,909	12.65
	8	4633	57,313	12.37
	9	6116	74,909	12.25
	10	6521	73,110	11.21
	11	6708	71,122	10.60
	12	6521	70,896	10.87

## 2.2 Time-series analysis of energy price variation

Time-series techniques have been employed in investigating the interrelationships across different energy markets. These studies, however, represent a separate research domain focusing on the stochasticity of energy markets without being incorporated into ESM. The targeted markets include crude oil, natural gas, and coal, among others, which are put in a relatively long-term scope to uncover their macro-economic impacts. A prevalent view is that oil prices are determined globally, while natural gas and coal prices are more determined regionally (Mohammadi, 2011). Although there is co-movement between the different markets in the long run, the price transmission mechanism is not the same, and in many cases, it presents an asymmetric and nonlinear manner (Atil et al., 2014). Under some market conditions, it is revealed that there exists short-run departures from the long-run equilibrium price between natural gas and crude oil, with many voices have noted that the price series appear to have decoupled (Ramberg & Parsons, 2012). It is also argued that technology is critical to the long-run relationship between fuel prices (Hartley et al., 2008). Moreover, significant cross-volatility spillover is found to exert persistent volatility effects for other markets such as agriculture and precious metals (Du et al., 2011; Kang et al., 2017).



**Fig. 1** The integrated modelling framework for linking interrelated energy price simulation with a stochastic energy system optimization model

## 2.3 Research goals

Energy market fluctuation is considered one of the most impactful factors in the energy system. To our knowledge, only a few studies are incorporating time-series analytical methods into energy system modeling (Kim et al., 2019; Lima et al., 2018), whereas a thorough and profound analysis is still absent regarding examining the mutual effect among multiple energy markets in ESM. Smaller-scale ESMs or single-period supply chain optimization models allow for implementation with methods such as Monte Carlo simulation or SP to deal with price uncertainty (Ren et al., 2020; Rezaee et al., 2017; Waltho et al., 2019). On the contrary, large-scale ESMs often build on relatively long timeframes; generating sufficient scenario trees for price over a long time horizon entails the curse of dimensionality. As a result, methods depending on distribution simulation of uncertain variables, such as stochastic programming, are computationally demanding and in many cases not even solvable. On the other hand, price scenarios with similar movement patterns in a large scenario set would actually result in the same outcomes of concern, indicating that reducing the number of price scenarios while retaining its crucial information of uncertainty is a feasible solution to overcome the computing burden. In this case, generating the most representative price scenarios becomes essential.

Figure 1 illustrates the structure of the proposed integrated modelling framework for linking interrelated energy price simulation with the ESM. This new approach employs time-series modelling of interrelated energy price variables and obtains a relatively small number of distinct price scenarios from an iterative machine-learning clustering procedure. These representative price scenarios are subsequently introduced into the constructed stochastic ESM to calculate the strategic and operational outcome variables considering intertwined energy price variations. This new framework could improve the calculation efficiency of traditional stochastic ESMs, which reply on simulating numerous scenarios regarding stochastic variables, and retain the critical information of uncertainty simultaneously.

The contribution reflected in the research goals of this study lies in three aspects. First, constructing an integrated modelling framework for a complicated energy system consisting of three competing technologies subject to multiple intertwined energy markets. Second, introducing machine-learning clustering into the framework to improve computation efficiency by generating the most representative price scenarios and while retaining the

varying price dynamics. Third, applying the modelling framework to advancing the understanding of how the price variation in these intertwined energy markets affects the configuration and operation of the whole energy supply chain system.

### 3 Problem description

This study aims to address the strategic and tactical decision-making problems of technology choice and operation planning for the supply chain of an entire energy system from upstream fuel supply to downstream plastic consumption, taking into account the interrelations of variations of different energy markets. Specifically, the energy supply system starts from three upstream technologies fueled by crude oil, natural gas, and coal to produce ethylene, which is the dominant raw material to make plastics. In the downstream sector, various demands of these products from four customers need to be satisfied.

Mutual relationships between the oil, gas, and coal market are considered in a manner that, stochasticity of price movement processes is simulated throughout the timespan of decisions, for which key parameters of these processes are estimated by statistical time series analytics with a long-term historical dataset. This section describes the entire energy system of interest and provides vital assumptions, parameters, and data employed in this study.

#### 3.1 The overall supply chain system

The plastic supply chain considered in this study involves feedstock supply from three fuel types, intermediate production (ethylene and polyethylene), manufacturing of two plastic products, and the end-use customers with various demands of the plastic products (Fig. 2). Ethylene is the dominant chemical feedstock for producing plastics worldwide. There are several approaches to produce ethylene, among which a conventional technology is oil-based. Three technological processes within the scope of this study are briefly introduced as follows:

- In the oil-based approach, an ethylene plant is typically integrated with an oil refinery. Naphtha, an intermediate product distilled from oil, usually serves as the feedstock. Naphtha is converted into olefins (compounds mainly made up of ethylene, propane, and others) through steam cracking.

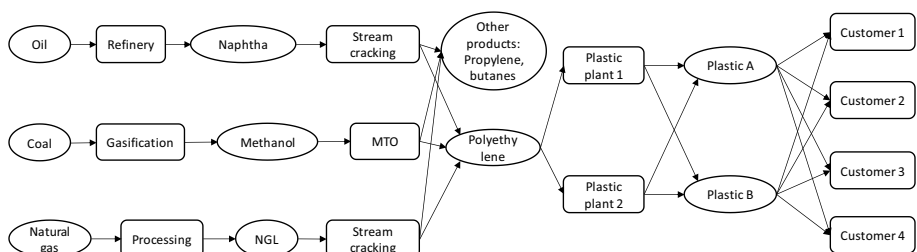


Fig. 2 Schematic of the overall plastic production supply chain

- The natural gas approach is favored in some regions with cheap, abundant natural gas resources. The shale gas revolution in the United States has caused over a decade-long downturn of the natural gas market, giving rise to the boom of producing ethylene in this approach. The general principle of the natural gas route is similar to the oil-based route; that is, steam cracking is also the main process to obtain the olefins. The difference lies in that the feedstock in the gas route is primarily a mixture of ethane and propane, normally extracted from the light components in processing raw natural gas.
- Coal and methanol-based technology emerged in China as a responsive measure to address the concern of energy security. Crude oil and natural gas resources are much scarcer in China, in contrast with coal abundance. As a result, various coal-based technologies have been developed and commercialized to produce liquid fuels and chemicals. This route uses coal to synthesize methanol, which serves as the feedstock to produce olefins through methanol-to-olefins (MTO).

Coal-based technology is contentious because it is carbon-intensive and loses economic competitiveness when oil price and nature price are low, as has been happening since the oil market slump in 2014. Even though this approach could ease the supply tension of oil and gas, the drawback is that its capital cost is extremely high compared to the other two alternatives, causing potential sunk costs. Therefore, a strategic analysis of the technological route choice is required.

The subsequent process is similar for each of the three technological routes; that is, ethylene is separated from the olefin mixture and then synthesized into polyethylene (PE). The solid PE resin is much easier for transportation. In real industrial practices, there are several types of PE, e.g., low-density polyethylene (LDPE) and high-density polyethylene (HDPE), among others. Numerous types of downstream plastics, in the forms of sheet, pipe, profile, etc., could be manufactured from PE resin. For the sake of brevity, we choose two types, namely, Plastic A and B, demanded by the four customers. Detailed specifications of technologies, cost parameters, installation setups, and other relevant data in the entire supply chain are provided in the next section.

### 3.2 Parameters and data

Techno-economics of the three technologies determines to a large extent the configurations of the supply chain. To lay a common ground for this comparison, we assume that the three technological routes have the same ethylene production capacity, which is 1.095 million tons (Mton) per annum. We compiled specific data for each cost item incurred in the technologies, see Table 2. However, we couldn't find data for all three technologies with the same capacity. As a result, we conduct a comprehensive analysis and calculate the capital cost from the reference case through the following formula:

$$CI = CI_{ref} \cdot \left( \frac{CAP}{CAP_{ref}} \right)^{\beta} \quad (1)$$

where  $CI$  is the capital investment,  $CI_{ref}$  is the capital investment of the reference case with the data compiled from literature (Xiang et al., 2014; Yang & You, 2017),  $CAP$  is the capacity scale,  $CAP_{ref}$  is the capacity of the reference case from literature,  $\beta$  is the scaling factor measuring the scale effect of cost.



It is noteworthy that there are many byproducts besides ethylene or propylene from the three technological processes. It is thus necessary to evaluate the costs solely attributable to ethylene. Different methods for cost allocation were proposed (Tayyari & Parsaei, 1992), including the mass-based measurement, economic value-based method, etc. Each of these methods has its application pre-conditions. For example, the mass-based method requires all co-products are of approximately the same level of economic value. Given that the considerable difference of economic value across the co-products in the case of this study, the economic value-based method is employed to determine a set of cost allocation factors by calculating the fraction of economic value for each co-product (Tayyari & Parsaei, 1992; Yang & You, 2017).

Key data and assumptions for the downstream installations, i.e., the plastic manufacturing plants and the customers, are summarized in Table 3. The initial demands of the four customers for the two products at year  $t = 1$  are also provided in the table. The demands are assumed to follow a general growth trend with a rate of 5% per annum, which randomly varies at the monthly level. As such, at an early stage, when the total demand is low, it does not need to open full capacities of all the plants. However, with demand gradually increasing, it would require more capacities to be added to the supply chain.

## 4 Models

### 4.1 System optimization

The problem of interest involves strategic and tactical decisions to be taken for a multi-echelon, multi-product, multi-period energy supply system. Several sources of uncertainty are considered with regard to different energy markets. The strategic decision deals with the selection of upstream technologies facing uncertain energy prices over the whole life span (25 years for all the three technologies), along which multiple energy markets are highly volatile and correlated. The tactical decision is to be undertaken in monthly operations, that is, the production scheduling among the three technologies given uncertain feed-stock prices, product prices, and demand variations. It, therefore, boils down to a stochastic optimization problem. As a result, the total amount of time steps for the decision-making is 300 months in 25 years. Planning of plastic production and dispatch is optimized to maximize the expected profit along the horizon of these sub-periods for all the price scenarios.

- Sets

$t \in T$ , period in the time horizon for the strategic problem  $T = \{1, \dots, LF_f\}$ , where  $f$  stands for a certain fuel-based technology.

$m \in M$ , month in one year, sub-annual time period for operations of the tactical problem,  $M = \{1, \dots, 12\}$ .

$f$ , index of fuel type or upstream technologies. As each of the three technologies corresponds with one fuel type, it therefore also represents technology type.  $f \in \{\text{oil}, \text{gas}, \text{coal}\}$ .

$p$ , index of plastic product type. It represents two types of products with different economic values.  $p \in \{A, B\}$ .

$pl$ , index of plastic manufacturing plant.  $pl \in \{pl1, pl2\}$ .

$c$ , index of customer,  $c \in \{c1, c2, c3, c4\}$ .

- $s$ , index of price scenario,  $s \in \{PS1, PS2, PS3, PS4, PS5, PS6\}$ .
- Parameters
    - $dr$ , discount rate, constant scalar parameter.
    - $ct_{f,t,m}^{up}$ , amortized capital investment for technology  $f$  at month  $m$  of year  $t$ , million yuan.
    - $cv_{f,t,m}^{up}$ , cost of variable O&M for technology  $f$  at month  $m$  of year  $t$ , million yuan.
    - $cf_{f,t,m}^{up}$ , cost of fixed O&M for technology  $f$  at month  $m$  of year  $t$ , million yuan.
    - $cr_{f,t,m}^{up}$ , consumption rate of fuel  $f$  at month  $m$  of year  $t$ , Mton.
    - $ct_{pl,t,m}^{ma}$ , amortized capital investment for manufacturing plant  $pl$  at month  $m$  of year  $t$ , million yuan.
    - $cv_{pl,t,m}^{ma}$ , cost of variable O&M for manufacturing plant  $pl$  at month  $m$  of year  $t$ , million yuan.
    - $cf_{pl,t,m}^{ma}$ , cost of fixed O&M for manufacturing plant  $pl$  at month  $m$  of year  $t$ , million yuan.
    - $conv_f^{up}$ , conversion rate of technology  $f$  from fuel  $f$  to ethylene.
    - $conv_{pl,p}^{ma}$ , conversion rate of manufacturing plant  $pl$  from ethylene to product  $p$ .
    - $er_f$ , CO<sub>2</sub> emission rate of technology  $f$ , ton CO<sub>2</sub>/ton.
    - $cap_f^{up}$ , production capacity of technology  $f$ , Mton/yr.
    - $cap_{pl,p}^{ma}$ , capacity of manufacturing plant  $pl$  for producing product  $p$ , Mton/yr.
    - $pb_s$ , probability of scenario  $s$ .
    - $dm_{c,p,t,m}$ , demand for plastic product  $p$  from customer  $c$  at month  $m$  of year  $t$ , Mton.
  - Stochastic parameters
    - $pr_{s,f,t,m}$ , price of fuel type  $f$  at month  $m$  of year  $t$  in scenario  $s$ , yuan/ton.
    - $pr_{s,p,t,m}$ , price of plastic product  $p$  at month  $m$  of year  $t$  in scenario  $s$ , yuan/ton.
  - Decision variables
    - $FD\_PURC_{s,f,t,m}^{up}$ , amount of feedstocks purchased by upstream technology  $f$  at month  $m$  of year  $t$  in scenario  $s$ , Mton.
    - $FD\_PURC_{s,pl,t,m}^{ma}$ , amount of ethylene purchased by manufacturing plant  $pl$  at month  $m$  of year  $t$  in scenario  $s$ , Mton.
    - $COM\_PROD_{s,f,t,m}^{up}$ , amount of ethylene produced from  $f$  at month  $m$  of year  $t$  in scenario  $s$ , Mton.
    - $COM\_PROD_{s,pl,p,t,m}^{ma}$ , amount of product  $p$  produced from manufacturing plant  $pl$  at month  $m$  of year  $t$  in scenario  $s$ , Mton.
    - $SEL_{s,f,t,m}^{up}$ , binary variable, when upstream technology  $f$  is selected at a certain point,  $SEL_{s,f,t,m}^{up} = 1$  for all the subsequent time-step at month  $m$  of year  $t$  in scenario  $s$ ,  $SEL_{s,f,t,m}^{up} = 0$  otherwise.
    - $SEL_{s,pl,t,m}^{ma}$ , binary variable, when manufacturing plant  $pl$  is selected at a certain point,  $SEL_{s,pl,t,m}^{ma} = 1$  for all the subsequent time-step at month  $m$  of year  $t$  in scenario  $s$ ,  $SEL_{s,pl,t,m}^{ma} = 0$  otherwise.
    - $OPEN_{s,f,t,m}^{up}$ , binary variable,  $OPEN_{s,f,t,m}^{up} = 1$  when upstream technology  $f$  is open at month  $m$  of year  $t$  in scenario  $s$ ,  $OPEN_{s,f,t,m}^{up} = 0$  otherwise.
    - $OPEN_{s,pl,t,m}^{ma}$ , binary variable,  $OPEN_{s,pl,t,m}^{ma} = 1$  when manufacturing plant  $pl$  is open at month  $m$  of year  $t$  in scenario  $s$ ,  $OPEN_{s,pl,t,m}^{ma} = 0$  otherwise.
  - Outcome variables
    - $REV_{p,t,m}$ , total revenue for sales of product  $p$  at month  $m$  of year  $t$ , million yuan.
    - $COST_{t,m}^{up}$ , total cost of upstream production at month  $m$  of year  $t$ , million yuan.
    - $COST_{t,m}^{ma}$ , total cost of the manufacturing stage at month  $m$  of year  $t$ , million yuan.
    - $EM_{t,m}$ , total CO<sub>2</sub> emissions at month  $m$  of year  $t$ , Mton.

The objective of the strategic decision-making problem is to maximize the total profit over the whole technology lifetime, as formulated in Eq. (2):

$$\max \text{ profit} = \text{REV} - \text{COST}^{\text{up}} - \text{COST}^{\text{ma}} \quad (2)$$

where the revenue  $\text{REV}_p$  from selling product  $p$ , the cost item  $\text{COST}^{\text{up}}$  for upstream production, and the other cost item  $\text{COST}^{\text{ma}}$  incurred at the manufacturing stage are given by Eqs. (3)–(5), and total CO<sub>2</sub> emissions are calculated by Eq. (6).

$$\begin{aligned} \text{REV} = & \sum_s \sum_{pl} \sum_p \sum_t \sum_m pb_s \cdot pr_{s,p,t,m} \cdot \text{COM\_PROD}_{s,pl,t,m}^{\text{ma}} / (1 + dr)^t \\ & + \sum_s \sum_f \sum_t \sum_m pb_s \cdot pr_{s,p,t,m} \cdot \text{COM\_PROD}_{s,f,t,m}^{\text{ma}} / (1 + dr)^t \end{aligned} \quad (3)$$

$$\begin{aligned} \text{COST}^{\text{up}} = & \sum_s \sum_f \sum_t \sum_m pb_s \cdot ci_{f,t,m}^{\text{up}} \cdot \text{SEL}_{s,f,t,m}^{\text{up}} / (1 + dr)^t \\ & + \sum_s \sum_f \sum_t \sum_m pb_s \cdot cf_{f,t,m}^{\text{up}} \cdot \text{OPEN}_{s,f,t,m}^{\text{up}} / (1 + dr)^t \\ & + \sum_s \sum_f \sum_t \sum_m pb_s \cdot \left( cv_{f,t,m}^{\text{up}} / \text{conv}_f^{\text{up}} + pr_{s,f,t,m} \right) \cdot \text{FD\_PURC}_{s,f,t,m}^{\text{up}} / (1 + dr)^t \end{aligned} \quad (4)$$

$$\begin{aligned} \text{COST}^{\text{ma}} = & \sum_s \sum_{pl} \sum_t \sum_m pb_s \cdot ci_{pl,t,m}^{\text{ma}} \cdot \text{SEL}_{s,pl,t,m}^{\text{ma}} / (1 + dr)^t \\ & + \sum_s \sum_{pl} \sum_t \sum_m pb_s \cdot cf_{pl,t,m}^{\text{ma}} \cdot \text{OPEN}_{s,pl,t,m}^{\text{ma}} / (1 + dr)^t \\ & + \sum_s \sum_{pl} \sum_t \sum_m pb_s \cdot \left( \sum_p cv_{f,t,m}^{\text{ma}} / \text{conv}_{pl,p}^{\text{ma}} + pr_{s,f,t,m} \right) \cdot \text{FD\_PURC}_{s,pl,t,m}^{\text{ma}} / (1 + dr)^t \end{aligned} \quad (5)$$

$$\text{EM} = \sum_s \sum_f pb_s \cdot \text{COM\_PROD}_{s,f,t,m}^{\text{up}} \cdot er_f \quad (6)$$

As shown by the equations, both the upstream and the manufacturing costs consist of capital investment, fixed and variable O&M costs at each subsequent time step. The main difference lies in fuel cost, which is determined by stochastic fuel prices and significantly affects the upstream production cost as well as the technological choice. The tactical problem defines monthly operations of the installations in the whole supply chain. Production planning deals with the timing and the amount of the products produced from each fuel-based technology and manufacturing plant at monthly time step accounting for the uncertain market prices in each scenario. The consequent CO<sub>2</sub> emissions are calculated by multiplying the amount of ethylene produced from each technology and the corresponding emissions rate. The summation of the emissions from each technology gives the total CO<sub>2</sub> emissions of the whole system, as shown in Eq. (6).

Material balances and physical constraints are ensured by Eqs. (7) and (16). Equation (7) restricts that volatile demand with a generally growing trend shall always be satisfied for each product at each time step. Equations (8)–(10) guarantee the material balances across all the echelons throughout the supply chain. Equations (11) and (12) indicate the capacity

constraints on both the upstream and the manufacturing plants. Equations (13)–(16) ensure that monthly operations of a particular technology rest on the prerequisite selection of this technology on a strategic level.

$$\sum_c dm_{c,p,t,m} = \sum_{pl} COM\_PROD_{s,pl,p,t,m}^{ma}, \quad \forall s, p, t, m \quad (7)$$

$$FD\_PURC_{s,pl,t,m}^{ma} = \sum_p COM\_PROD_{s,pl,p,t,m}^{ma} / CONV_{pl,p}^{ma} \quad \forall s, pl, t, m \quad (8)$$

$$FD\_PURC_{s,f,t,m}^{up} = COM\_PROD_{s,f,p,t,m}^{up} / CONV_f^{up} \quad \forall s, f, t, m \quad (9)$$

$$\sum_{pl} FD\_PURC_{s,pl,t,m}^{ma} = \sum_f COM\_PROD_{s,f,t,m}^{up} \quad \forall s, t, m \quad (10)$$

$$COM\_PROD_{s,f,t,m}^{up} \leq cap_f^{up} \cdot OPEN_{s,f,t,m}^{up} \quad \forall s, f, t, m \quad (11)$$

$$COM\_PROD_{s,pl,p,t,m}^{ma} \leq cap_{pl,p}^{ma} \cdot OPEN_{s,pl,p,t,m}^{ma} \quad \forall s, pl, p, t, m \quad (12)$$

$$OPEN_{s,f,t,m}^{up} \leq SEL_{s,f,t,m}^{up} \quad \forall s, f, t, m \quad (13)$$

$$OPEN_{s,pl,t,m}^{ma} \leq SEL_{s,pl,t,m}^{ma} \quad \forall s, pl, t, m \quad (14)$$

$$SEL_{s,f,t-1,m}^{up} \leq SEL_{s,f,t,m}^{up} \quad \forall s, f, t, m \quad (15)$$

$$SEL_{s,f,t-1,m}^{ma} \leq SEL_{s,f,t,m}^{ma} \quad \forall s, pl, t, m \quad (16)$$

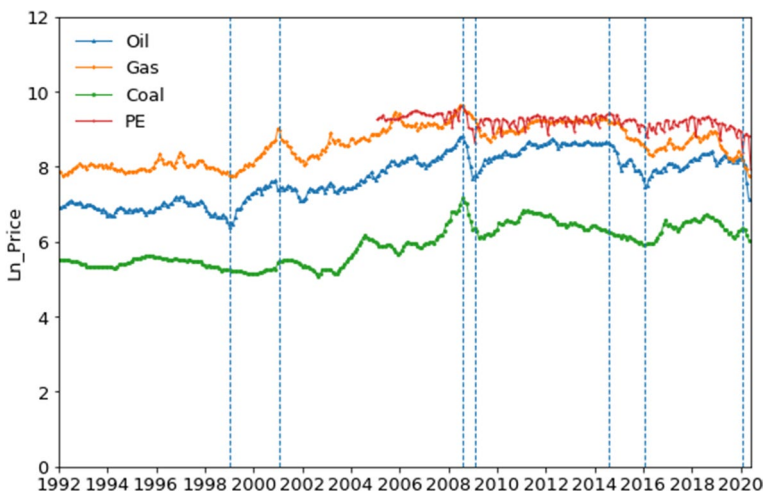
## 4.2 Modelling price variation and interrelations

Studies reveal that changes in energy prices could be transmitted horizontally and/or vertically, in which horizontal transmission is defined as changes generated by linkages among fuels at a similar stage of processing, such as from crude oil to natural gas, whereas vertical transmissions pertain to linkages between upstream/downstream in the energy supply chain, e.g., from oil to petrochemical products (Kaufmann et al., 2009). Numerous studies have been carried out using various methods of statistics to uncover these mutual effects. Considerable difference exists with regards to analytical methods, targeted markets, data sources, and time spans, leading to diverged interpretations of these interrelations. Despite this, some common findings have been drawn. For instance, oil product prices (such as gasoline) do not react symmetrically to changes in crude oil prices. This asymmetric effect is also referred to as ‘rockets and feathers’, a metaphor indicating that downstream product prices rise faster in response to an increase in the price of crude oil than prices decline in response to a drop in crude oil prices (Kaufmann and Laskowski, 2005; Peltzman, 2000; Tappata, 2009). Moreover, long-run cointegrating relationships at the horizontal level have

been tested by studies using Vector Autoregression (VAR) or Vector Error Correction Model (VECM), with conclusions that innovations in crude oil prices ‘Granger cause’ natural gas prices, but not in a reversed way (Kaufmann et al., 2009). The coal market is usually deemed local market and not significantly affected by other international markets such as crude oil. This has been revealed by previous statistic studies as discussed above.

A study on energy markets in the United States finds that local coal and natural gas prices as well as coal and crude oil, are only very superficially linked (Bachmeier & Griffin, 2006). In China’s case, it is found that China’s coal and crude oil prices largely hinge on the shares of oil and coal in China’s energy mix, and inter-fuel substitution dominated the interaction of China’s coal market with other energy types (Li et al., 2019). It might be because, in China, the majority of coal trading is in domestic markets; therefore, fluctuations in international oil and gas markets exert relatively weak impacts.

We compiled the historical prices of crude oil, natural gas, coal, and PE for this study. The logarithmic values of these prices are shown in Fig. 3. There are different prices for one commodity as many trading markets exist; we tried to collect the most representative price indices as far as possible. For example, crude oil is regarded as the most internationally traded commodity; we thus use the price indices compiled by the International Monetary Fund (IMF), which takes the average values of several benchmark futures prices, including WTI, Brent, and Dubai (IMF, 2020). The price indicators for natural gas and coal in this study are also collected from this dataset. As for the PE prices, we use the data from Wind Data Service (2020), which covers a relatively shorter time span dating back to 2005. All the original price indicators are converted to a uniformed unit and then taken logarithm to allow comparison in the same scale. The vertical blue lines in the figure signify the co-movements among these commodities in the periods of several historical events. The quick rise of the energy prices from 1999 was mainly attributed to the increase in energy demand in developing countries such as China and India. In the middle of the 2008 financial crisis, the prices of oil, natural gas, and coal underwent a significant decrease, along with many other commodities. After recovered from the financial crisis, energy prices stagnated until the plummet of oil price in the middle of 2014. This was mainly driven by the compound



**Fig. 3** Monthly logarithmic prices of crude oil, natural gas, coal and PE

effects of a slowing global economy, positive supply shocks to oil production, among others (Baumeister & Kilian, 2016). The latest COVID-19 pandemic lowered energy demand because of lockdowns around the world, which also significantly hit the three energy markets as well as other financial markets.

#### 4.2.1 Oil price scenarios

According to the statistic findings drawn from the above-mentioned studies, we simulate the prices of these commodities based on the assumptions that oil price follows an independent stochastic process, which also exerts influences on other price movements. As a result, there are in total four stochastic price processes in this model. The Ornstein–Uhlenbeck model is employed to simulate the movement of oil price, given by Eq. (17)

$$d \ln pr_{oil,t} = \lambda(\mu - \ln pr_{oil,t})dt + \sigma dW_t \quad (17)$$

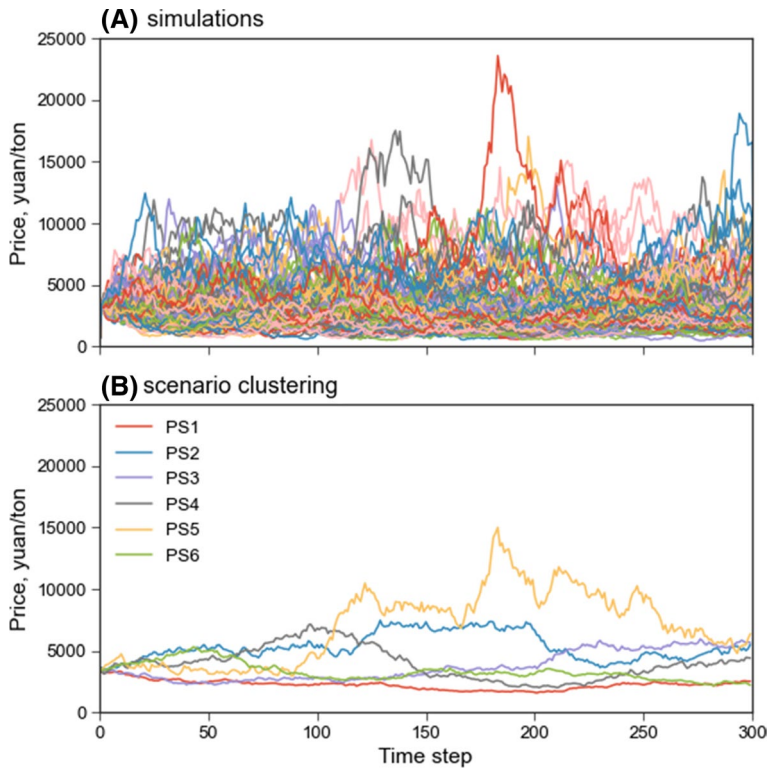
where  $\lambda$  is the mean-reversion rate,  $\mu$  is the mean,  $\sigma$  is the volatility, and  $W_t$  is the Wiener process. The Ornstein–Uhlenbeck process features a ‘mean-reverting’ trend in a way that the drift term becomes positive when the current value of the process is less than the long-term equilibrium level for the process represented by  $\mu$ ; otherwise, the drift term turns negative. This creates a tendency of the process to return to its long-term equilibrium, representing the characteristic of the historical movement of oil price. The parameters for this process are estimated through historical data. Simulations of price movement projected in 300 time-steps are performed. Panel A of Fig. 4 shows 100 stochastic paths of simulation. Each path represents a price scenario (PS) and serves as inputs to the optimization model.

It is important to note that including all the simulated price paths would pose significant pressure on computing resources, sometimes even fail to attain feasible results within reasonable time. In addition, it is not necessary to incorporate all the price paths because many of the paths are close to each other, and therefore would not significantly impact the strategies obtained from the optimization model. In fact, the obtained optimal strategies predominantly rely on each distinctive ‘price pattern’ that can be extracted from the simulated paths, for example, the magnitude of price variation and the timing of abrupt changes. Therefore we use the  $k$ -means clustering algorithm to partition  $n = 100$  simulated price paths into a few price scenarios, in which each path belongs to the cluster with the nearest mean  $\delta$ . The objective function is given by Eq. (18).

$$\arg \min_{PS} \sum_{i=1}^k \sum_{pr \in PS_i} \|x - \delta_i\| = \arg \min_{PS} \sum_{i=1}^k |PS_i| VarPS_i \quad (18)$$

where  $pr$  is the simulated price path,  $PS$  is the price scenario to be clustered.

The number of clusters is determined through a procedure termed cluster validation aiming at the quantitative evaluation of the results of the clustering algorithms. There are different methods for cluster validation (Rendón et al., 2011). Here we adopt compactness (CP) and separation (SP) to measure the internal and external clustering results, respectively, and use the ratio of the two indices as the criterion to select the number of clusters. Calculation of CP and SP is given by Eqs. (19) and (20), see Table 4. According to this criterion, the number of clusters is determined as six.



**Fig. 4** Oil price simulations and scenario clustering

$$\overline{CP}_i = \frac{1}{|\Omega_i|} \sum_{x \in \Omega_i} \|x_i - w_i\|, \quad \overline{CP} = \frac{1}{K} \sum_{k=1}^K \overline{CP}_k \quad (19)$$

$$\overline{SP} = \frac{2}{k^2 - k} \sum_{i=1}^k \sum_{j=s+1}^k \|w_i - w_j\| \quad (20)$$

where  $K$  is the number of clusters, and  $w$  is the centroid of cluster  $k$ .

The six scenarios obtained from clustering are illustrated in Panel B of Fig. 4. The clustered price scenarios demonstrate distinctive patterns from each other. For example, PS2 presents a relatively stable movement, in contrast with PS4, which shows significant variation in the time frame. Nevertheless, all the price paths follow a ‘mean-reverting’ trend, as they are all governed by the same Ornstein–Uhlenbeck process.

#### 4.2.2 Natural gas price and plastic price

The above-mentioned relation between the oil market, natural gas market, and downstream product market indicates that modeling natural gas price and plastic product price shall

account for the impacts from the oil price. As a result, we use Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX) model to project natural gas price and product price, in which oil price serves as the exogenous variable.

$$y_{g,t} = \beta y_{o,t} + \phi_1 y_{g,t-1} + \cdots + \phi_p y_{g,t-p} - \phi_1 z_{g,t-1} - \cdots - \phi_q z_{g,t-q} + z_{g,t} \quad (21)$$

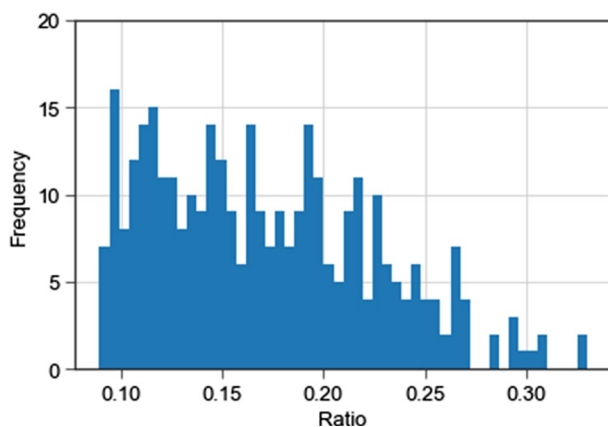
where  $y_{g,t}$  and  $y_{o,t}$  are the integrated time series of natural gas price and oil price,  $z_{g,t}$  is a white noise process. Note that  $y_{g,t}$  and  $y_{o,t}$  are supposed to be stationary time series. To achieve this, a conventional way is to take  $d$ -order difference until the resulted time series becomes stationary. This method is suitable for forecasting when multivariate has any type of data pattern, for instance, with trend or seasonality or cyclicity. Specification of the model should follow a particular procedure, which is elaborated in the supplementary information (SI). The values of the auto-correlation and partial auto-correlation function of any series with its lagged values are provided by auto-correlation function (ACF) and partial auto-correlation function (PACF) shown in Figs. S3 (natural gas) and S4 (plastic product), respectively. Following this procedure, it is specified that  $p = 2$ ,  $q = 2$  for the natural gas model, and  $p = 1$ ,  $q = 1$  for the plastic price model. Durbin Watson (DW) test was also performed to examine the distributions of residuals (Figs. S5 and S6). The modelled price projections are presented below.

### 4.2.3 Coal price

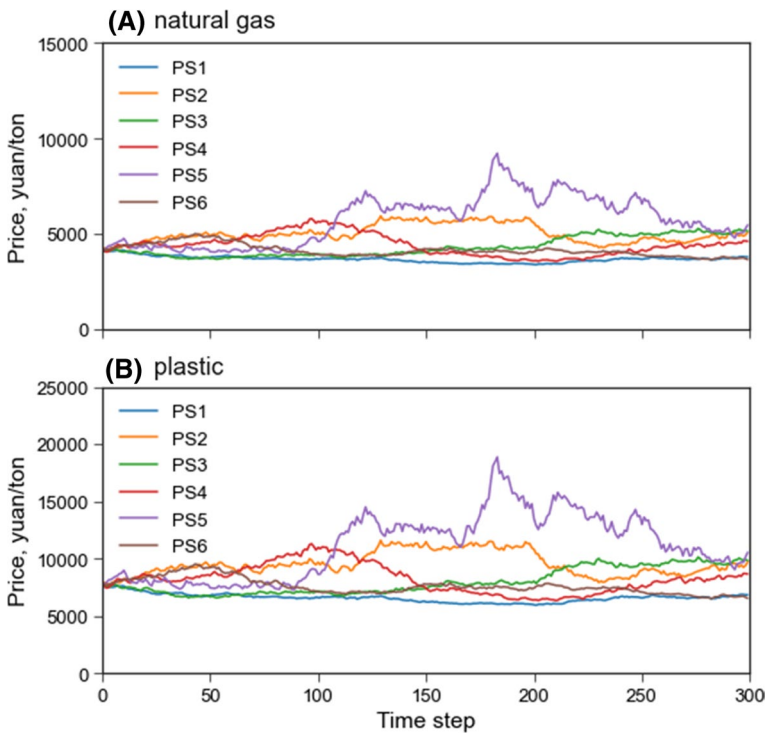
As analyzed previously, the interrelation between coal and oil differs from the one between oil and gas. The most distinct co-movements between coal and other energy commodities occur in several business cycles, which affect many commodity prices not limited to energy markets. Analyzing the effects of business cycles is beyond the scope of this study. For the sake of brevity, we use Monte Carlo simulations to capture this interrelation. The ratio of coal price relative to oil price is set as a random variable, and its value is sampled at each time step from historical distribution through Monte Carlo simulation (Fig. 5). As a result, coal price is calculated by multiplying oil price and this random ratio.

Price projections for the four commodities in the six price scenarios are summarized in Fig. 6. Note that all the prices are taken logarithmic values in order to be set on the same scale. The interrelations of these four commodities in each scenario are illustrated. Specifically, PS5 is the most volatile and reaches the spikes in the middle of the time horizon.

**Fig. 5** Histogram of the ratio of coal price relative to oil price







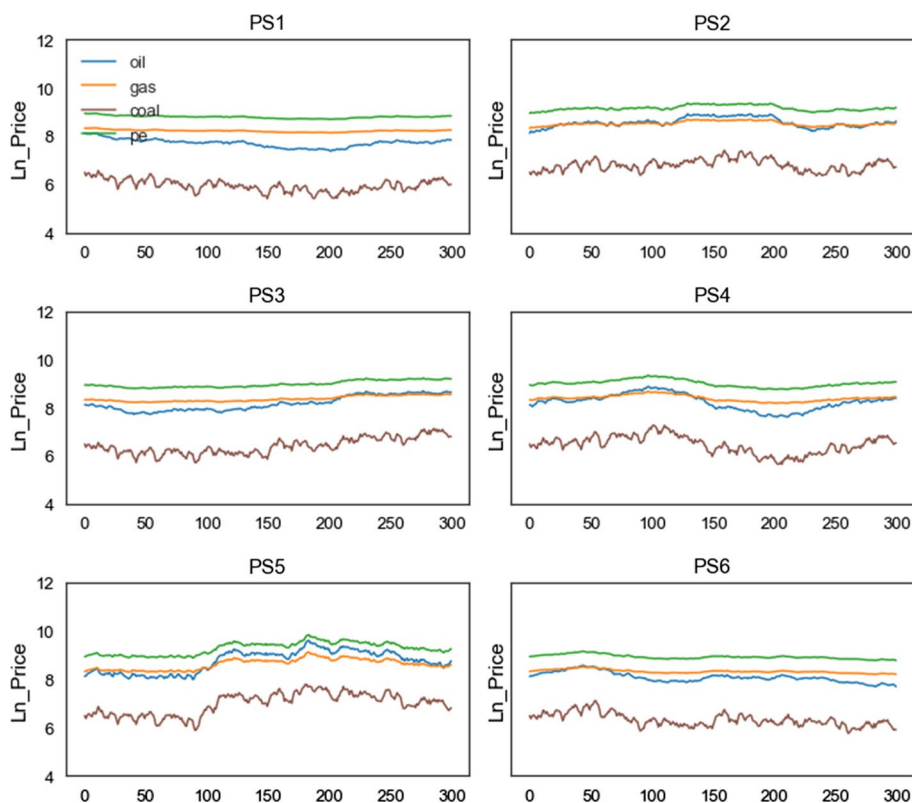
**Fig. 6** Modelling natural gas price and plastic price in the six clustered scenarios

In contrast, PS1 is the most flattened with very small variations. Others are in between. This reflects the representability of these scenarios in terms of price pattern distinctiveness. The blue lines represent oil price, modelled by the Ornstein–Uhlenbeck process with information on its own historical data. Prices of natural gas and plastic, as elaborated in the previous sections, are modelled by the ARIMAX model in which oil price is adopted as an exogenous variable to account for the external impacts. The simulated coal prices, despite following a different modelling approach, also capture the co-movement tendency in the overall energy markets (Fig. 7).

### 4.3 Model implementation

All the components in the model are coded in Python 3.7. The core optimization component is constructed in Pyomo 5.6, a Python-based optimization modeling framework, with Gurobi 8.1 serving as the solver. Modelling price variations are mainly implemented in the Python-based statistical module Statsmodels 0.11. The python-based machine learning package, Scikit-learn 0.23, is employed to perform k-means clustering of price scenarios.

As explained above, the proposed integrated modelling framework consists of an ESM for optimizing the energy supply system and several stochastic process models for generating a variety of energy price scenarios. The ESM comprises approximately 6251 variables



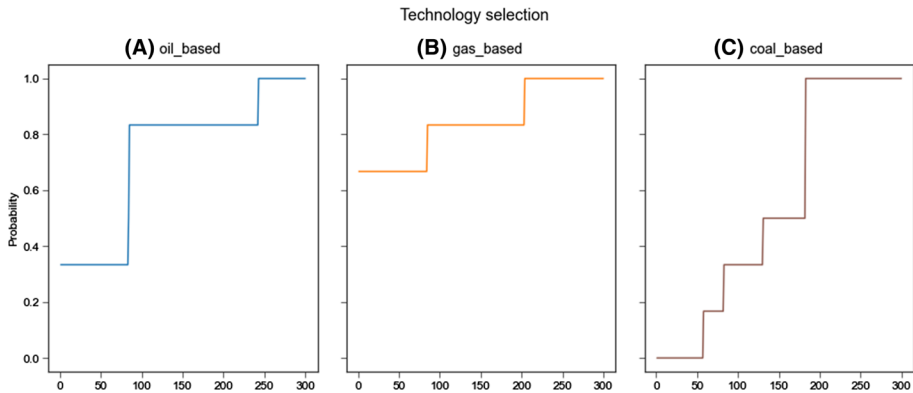
**Fig. 7** Logarithmic prices of the four energy commodities in the six scenarios

and 8423 constraints. It takes roughly 1800s for running one price scenario, and 11,000 s for the six scenarios in total. The computing platform includes Intel Core i7-6500U CPU and 32 GB RAM, with Windows 10 as the operating system.

## 5 Results and analysis

### 5.1 Technology selection and capacity expansion

Technology selection is mainly determined by the relative economic advantages and the varying interrelations of the various energy prices. At the early stage, the total demand for plastic products is at a lower level, and the capacity of one single technology is sufficient to meet the demand. As the demand grows, more capacities would be required, and therefore other technologies need to be added at some points. Figure 8 shows the probabilities of the three technologies being selected along the horizon. Note that all six price scenarios are assigned with the same probability. At the initial stage, the gas-based and oil-based approaches are selected with different probabilities. In contrast, the coal route is abandoned because of its economic disadvantage. This advantage of the oil and gas-based technologies with higher selection probabilities continues until time step 190. It is interesting to



**Fig. 8** Technology selection probability along the time horizon

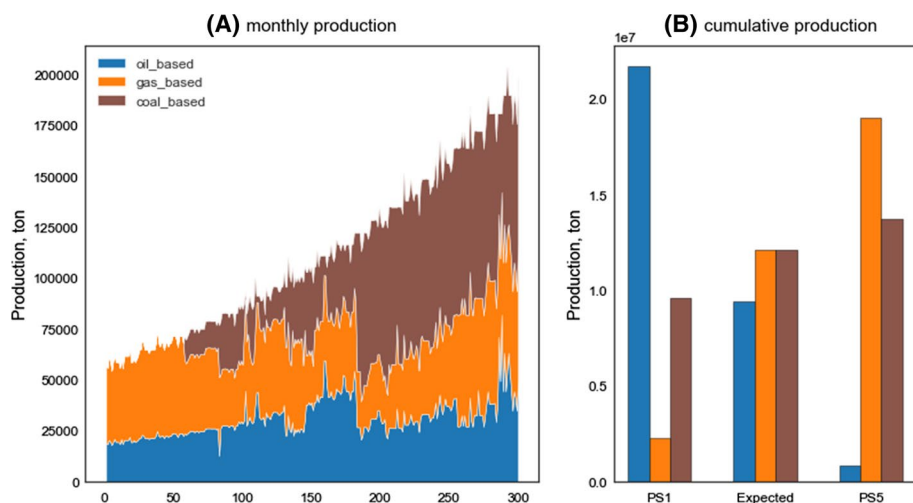
observe that coal-based technology is increasingly favored from time step 52 onwards. All three technologies are selected at the final stage because of the growing downstream demand. It is noteworthy that the strategic and tactic decision-making is mainly dependent on feedstock price variation because the feedstock cost constitutes the largest part in total costs.

## 5.2 Operational production scheduling

The operational strategy shows a more complicated manner, because it needs not only to decide which technology should be selected and switched on at each time step, but also how many products should be produced from each technology. These results are presented in Fig. 9. There are several time windows when frequent switching of the operational status occurs, for instance, the natural gas plant during time step 80–90 and 260–270. As coal price is gaining the economic advantage, it is gradually taking the shares of natural gas and oil in the second half of the horizon. This also reflects the competitiveness of the coal-based route featuring low fuel costs despite the highest capital investment among the three approaches. In the actual plant operations, such costs could be high, particularly in the process industry. One caveat of this modelling work is that the switching costs are not accounted for. As a result, the operational costs would be underestimated.

The operational planning of monthly ethylene production for the three technologies (Panel A) is in line with the results of technology selection. The natural gas route is more favored at the beginning, whereas the coal-based route is catching up after time step 52. All three technologies are selected and become operational after this initial stage. Nevertheless, the outputs from each plant are highly volatile. It is seen an abrupt drop for gas-based production at around time step 180 because of the surge of the natural gas price at that point.

The cumulative production (Panel B) could better illustrate the whole profile over the time span, where the two extreme cases, PS1 and PS5 are compared. The results of a deterministic version of the model for the six price scenarios are also calculated (details are summarized in the SI) and compared with the two scenarios. As discussed above, PS1 features a relatively stable, low-level price variation, which significantly



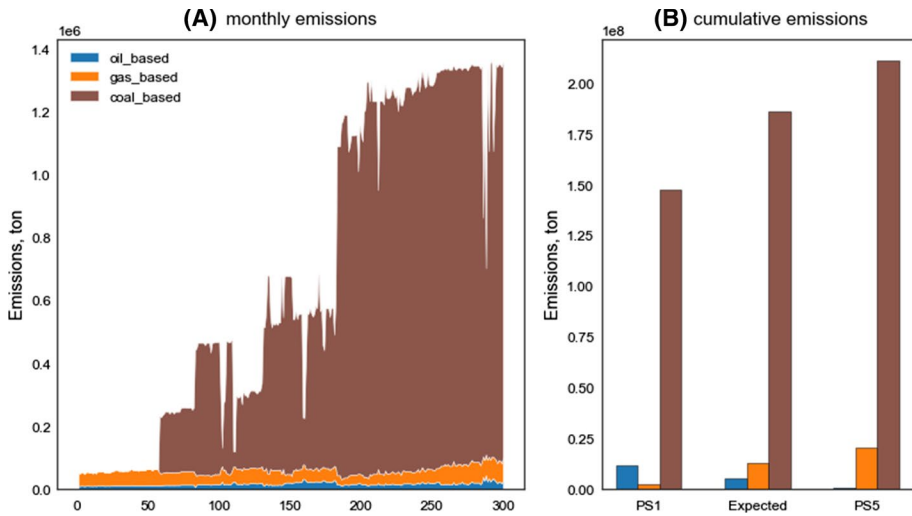
**Fig. 9** Monthly and cumulative production of ethylene in the operational planning

promotes the competitiveness of the oil-based route. On the contrary, the oil-based route gradually loses its market share to the gas-based and coal-based technologies in PS5. It is because when oil prices fluctuating at a relatively high level, the gaps between oil prices and the other two tend to enlarge. The expected results from the stochastic version show that the three technologies produce similar amounts of ethylene. This comparison of the results suggests that high volatilities of energy markets tend to shrink the competitiveness of the producing technologies. In addition, the frequent switching of the operation status, as shown in the second half of the duration in PS5, would inevitably lead to the reluctance of investors' decision-making and consequently hamper the development of the industry.

## 6 CO<sub>2</sub> emissions

CO<sub>2</sub> emissions from the operations of the three technologies are also calculated. Figure 10 presents the expected results for the stochastic model compared to the two extreme scenarios. The results show that the most carbon-intensive coal-based route dominates the total emissions of the three routes. Cumulative emissions from the coal-based technology constitute roughly 91% of the aggregated cumulative emissions of the three technologies, with only a very slight difference of the proportion within the three cases. However, the total emissions differ remarkably across the three cases, depending on how much coal is used. PS1 has the lowest emissions because of the dominance of the oil route in this scenario. In PS5, however, coal utilization is at a relatively higher level, causing a large share of emissions despite the gas-based technology produces most of the products.

To examine the impacts of carbon tax on the total system cost, four levels of carbon tax ranging from 50 to 200 yuan/ton were set. For each tax level, the ratio of the additional costs from carbon tax relative to the total system cost was calculated for each scenario.



**Fig. 10** CO<sub>2</sub> emissions of the three technologies

**Table 5** The impacts of carbon tax on the total system cost

Carbon tax (yuan/ton)	PS1 (%)	PS2 (%)	PS3 (%)	PS4 (%)	PS5 (%)	PS6 (%)	Expectation (%)
50	3.91	2.52	5.18	3.09	3.05	4.50	3.61
100	7.82	5.04	10.36	6.19	6.10	9.00	7.22
150	11.72	7.56	15.54	9.28	9.15	13.50	10.83
200	15.63	10.09	20.72	12.37	12.20	18.00	14.45

The results are summarized in Table 5. It is shown that PS3 has the highest cost increase, because this scenario has the highest emissions due to most usage of the coal-based route. It is also interesting to find that, PS1 and PS2, two scenarios in which emissions are very close, have distinct cost increase ratios when exposed to the same level of the carbon tax. It is because PS1 uses more coal than PS2, but the latter uses significantly larger amount of natural gas, which drives up the total system cost.

It is important to note that the potential costs of carbon emissions are not incorporated into the decision-making process in the model; hence the technological choice is not affected by the respective emission intensities of these technologies. Although not included in this study, this direction is definitely worthy of further investigation. Many countries are now planning to implement an emissions trading scheme (ETS) to build an economy-wide carbon trading market. Either ETS or carbon tax will put considerable costs on carbon-intensive industries, especially for coal-based technologies. Considering this factor would drastically change the technology portfolio.

## 7 Conclusions

Different energy markets are interrelated in a manner that the fluctuation in one market may transmit either horizontally to the parallel competing market or vertically to the downstream products. Investing technology among various competing options is determined by its own economic strength, i.e., the ability to maximize the total profit over the timespan. Although the co-movements of several energy commodity prices are typically observed both in the short- and long run, these interrelations vary significantly along the horizon. These varying interrelations greatly affect the competitiveness of energy technologies, therefore complicate the decision-making problem for a complex energy system characterized by multiple competing technologies, multiple echelons and multiple periods. This study investigates this problem by constructing an energy system optimization model and incorporating the dynamics of multiple energy prices with stochastic process simulations. Specifically, the oil price is modelled with the mean-reverting Ornstein–Uhlenbeck process, the key parameters of which are regressed by the historical data. Gas price and plastic price are modelled by the ARIMAX model, in which oil price is incorporated as an exogenous variable. Monte Carlo simulations are used to model coal prices. A large number of price simulations dramatically affects the efficiency of model computing. By clustering the numerous possible price paths into a few distinctive patterns, the model could be solved within reasonable time, while retaining key information from the results. Integrating machine learning clustering methods into the stochastic optimization model, as presented in this study, demonstrate potential for further application to other problems of this kind characterized by high-dimensional stochastic variables.

The model results suggest that high volatilities of energy markets compromise the competitiveness of the producing technologies and consequently cause reluctance of investment behavior, which would eventually barricade the industry development. It is shown that the coal-based route is favored under the scenario with extremely volatile oil prices. Nevertheless, the cost from CO<sub>2</sub> emissions is not included in this optimization model. The variation of emissions allowance market would inevitably add new risks on the related investment, hence influence the decision-making process, particularly considering the fact that the national emissions trading scheme is set to operate soon. In this context, reducing the volatilities of energy prices through a variety of measures becomes essential to create a sound market environment and to manage the market expectation of investors.

This study evaluates the interrelations across energy prices in a market consisting of several existing technologies without considering entering of new technologies. In fact, technological breakthrough may permanently break these market price interrelations (Hartley et al., 2008). There are some promising technologies, e.g., producing chemicals directly from crude oil (Tullo, 2019) or the widely debated biomass-based route. It is imperative to evaluate the possible technological learning trends of these innovations on the energy market once they are commercialized. Another aspect that deserves further exploration lies in assessing the impacts from the environmental perspective, including carbon emissions, water withdrawal, etc. For instance, high carbon-intensity of the coal-based technology would undermine its economic advantage when facing carbon pricing mechanisms such as ETS or carbon tax. Quantitative analysis in this regard shall be explored in future research.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10479-021-04229-3>.

**Acknowledgements** This research was supported by the Major Innovation & Planning Interdisciplinary Platform for the “Double-First Class” Initiative, Renmin University of China, the National Natural Science Foundation of China (71961137012, 71874055), and the International Cooperation Program of PetroChina (2018D-5009-06).

## References

- Alonso-Ayuso, A., Escudero, L. F., Guignard, M., & Weintraub, A. (2018). Risk management for forestry planning under uncertainty in demand and prices. *European Journal of Operational Research*, 267, 1051–1074.
- Atil, A., Lahiani, A., & Nguyen, D. K. (2014). Asymmetric and nonlinear pass-through of crude oil prices to gasoline and natural gas prices. *Energy Policy*, 65, 567–573. <https://doi.org/10.1016/j.enpol.2013.09.064>
- Awudu, I., & Zhang, J. (2013). Stochastic production planning for a biofuel supply chain under demand and price uncertainties. *Applied Energy*, 103, 189–196.
- Babonneau, F., & Haurie, A. (2019). Energy technology environment model with smart grid and robust nodal electricity prices. *Annals of Operations Research*, 274, 101–117. <https://doi.org/10.1007/s10479-018-2920-1>
- Bachmeier, L. J., & Griffin, J. M. (2006). Testing for market integration: Crude oil, coal, and natural gas. *The Energy Journal*, 27, 55–72.
- Baumeister, C., & Kilian, L. (2016). Understanding the decline in the price of oil since June 2014. *Journal of the Association of Environmental and Resource Economists*, 3, 131–158. <https://doi.org/10.1086/684160>
- Borodin, V., Bourtembourg, J., Hnaïen, F., & Labadie, N. (2016). Handling uncertainty in agricultural supply chain management: A state of the art. *European Journal of Operational Research*, 254, 348–359. <https://doi.org/10.1016/j.ejor.2016.03.057>
- BP, 2020. Statistical Review of World Energy 2020.
- Cunico, M. L., Flores, J. R., & Vecchiotti, A. (2017). Investment in the energy sector: An optimization model that contemplates several uncertain parameters. *Energy*, 138, 831–845. <https://doi.org/10.1016/j.energy.2017.07.103>
- Du, X., Yu, C. L., & Hayes, D. J. (2011). Speculation and volatility spillover in the crude oil and agricultural commodity markets: A Bayesian analysis. *Energy Economics*, 33, 497–503. <https://doi.org/10.1016/j.eneco.2010.12.015>
- Gorissen, B. L., Yanıkoğlu, İ., & den Hertog, D. (2015). A practical guide to robust optimization. *Omega*, 53, 124–137. <https://doi.org/10.1016/j.omega.2014.12.006>
- Grossmann, I.E., Apap, R.M., Calfa, B.A., García-Herreros, P., Zhang, Q., 2016. Recent advances in mathematical programming techniques for the optimization of process systems under uncertainty. In *Computers and chemical engineering, 12th international symposium on process systems engineering and 25th European symposium of computer aided process engineering (PSE-2015/ESCAPE-25)*, 31 May–4 June 2015, Copenhagen, Denmark (vol. 91, pp. 3–14). <https://doi.org/10.1016/j.compchemeng.2016.03.002>
- Hartley, P. R., Medlock, K. B., III., & Rosthal, J. E. (2008). The relationship of natural gas to oil prices. *The Energy Journal*, 29, 47–66.
- IMF, (2020). IMF Primary Commodity Prices.
- Kang, S. H., McIver, R., & Yoon, S.-M. (2017). Dynamic spillover effects among crude oil, precious metal, and agricultural commodity futures markets. *Energy Economics*, 62, 19–32. <https://doi.org/10.1016/j.eneco.2016.12.011>
- Kaufmann, R.K., Dees, S., & Mann, M. (2009). Horizontal and vertical transmissions in the US oil supply chain. *Energy Policy*, 37, 644–650. <https://doi.org/10.1016/j.enpol.2008.09.052>
- Kim, H., Cheon, H., Ahn, Y.-H., & Choi, D. G. (2019). Uncertainty quantification and scenario generation of future solar photovoltaic price for use in energy system models. *Energy*, 168, 370–379. <https://doi.org/10.1016/j.energy.2018.11.075>
- Kim, T. T., & Poor, H. V. (2011). Scheduling Power Consumption With Price Uncertainty. *IEEE Transactions on Smart Grid*, 2, 519–527. <https://doi.org/10.1109/TSG.2011.2159279>
- Li, J., Xie, C., & Long, H. (2019). The roles of inter-fuel substitution and inter-market contagion in driving energy prices: Evidences from China’s coal market. *Energy Economics*, 84, 104525. <https://doi.org/10.1016/j.eneco.2019.104525>



- Lima, C., Relvas, S., & Barbosa-Póvoa, A. (2018). Stochastic programming approach for the optimal tactical planning of the downstream oil supply chain. *Computers & Chemical Engineering*, 108, 314–336. <https://doi.org/10.1016/j.compchemeng.2017.09.012>
- Lund, H., Sorknæs, P., Mathiesen, B. V., & Hansen, K. (2018). Beyond sensitivity analysis: A methodology to handle fuel and electricity prices when designing energy scenarios. *Energy Research & Social Science*, 39, 108–116. <https://doi.org/10.1016/j.erss.2017.11.013>
- McCollum, D. L., Jewell, J., Krey, V., Bazilian, M., Fay, M., & Riahi, K. (2016). Quantifying uncertainties influencing the long-term impacts of oil prices on energy markets and carbon emissions. *Nature Energy*, 1, 1–8. <https://doi.org/10.1038/nenergy.2016.77>
- Mohammadi, H. (2011). Long-run relations and short-run dynamics among coal, natural gas and oil prices. *Applied Economics*, 43, 129–137. <https://doi.org/10.1080/00036840802446606>
- Ning, C., & You, F. (2019). Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Computers & Chemical Engineering*, 125, 434–448. <https://doi.org/10.1016/j.compchemeng.2019.03.034>
- Peltzman, S. (2000). Prices rise faster than they fall. *Journal of Political Economy*, 108, 466–502. <https://doi.org/10.1086/262126>
- Pfenninger, S., Hawkes, A., & Keirstead, J. (2014). Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews*, 33, 74–86. <https://doi.org/10.1016/j.rser.2014.02.003>
- Ramberg, D. J., & Parsons, J. E. (2012). The weak tie between natural gas and oil prices. *The Energy Journal*, 33, 13–36.
- Ren, H., Zhou, W., Guo, Y., Huang, L., Liu, Y., Yu, Y., Hong, L., & Ma, T. (2020). A GIS-based green supply chain model for assessing the effects of carbon price uncertainty on plastic recycling. *International Journal of Production Research*, 58, 1705–1723. <https://doi.org/10.1080/00207543.2019.1693656>
- Ren, H., Zhou, W., Makowski, M., Yan, H., Yu, Y., & Ma, T. (2019). Incorporation of life cycle emissions and carbon price uncertainty into the supply chain network management of PVC production. *Annals of Operations Research*, 82, 1–20. <https://doi.org/10.1007/s10479-019-03365-1>
- Rendón, E., Abundez, I., Arizmendi, A., & Quiroz, E. M. (2011). Internal versus external cluster validation indexes. *International Journal of Computers and Communications*, 5, 27–34.
- Rezaee, A., Dehghanian, F., Fahimnia, B., & Beamon, B. (2017). Green supply chain network design with stochastic demand and carbon price. *Annals of Operations Research*, 250, 463–485.
- Rout, U. K., Akimoto, K., Sano, F., Oda, J., Homma, T., & Tomoda, T. (2008). Impact assessment of the increase in fossil fuel prices on the global energy system, with and without CO<sub>2</sub> concentration stabilization. *Energy Policy*, 36, 3477–3484. <https://doi.org/10.1016/j.enpol.2008.05.030>
- Seljom, P., & Tomasgard, A. (2015). Short-term uncertainty in long-term energy system models—A case study of wind power in Denmark. *Energy Economics*, 49, 157–167. <https://doi.org/10.1016/j.eneco.2015.02.004>
- Soroudi, A., Siano, P., & Keane, A. (2016). Optimal DR and ESS scheduling for distribution losses payments minimization under electricity price uncertainty. *IEEE Transactions on Smart Grid*, 7, 261–272. <https://doi.org/10.1109/TSG.2015.2453017>
- Tappata, M. (2009). Rockets and feathers: Understanding asymmetric pricing. *The RAND Journal of Economics*, 40, 673–687. <https://doi.org/10.1111/j.1756-2171.2009.00084.x>
- Tayyari, F., & Parsaei, H. R. (1992). Joint cost allocation to multiple products: Cost accounting v. engineering techniques. In H. R. Parsaei & A. Mital (Eds.), *Economics of advanced manufacturing systems* (pp. 189–199). Springer.
- Uryasev, S., 2000. Conditional value-at-risk: Optimization algorithms and applications. In *Presented at the proceedings of the IEEE/IAFE/INFORMS 2000 conference on computational intelligence for financial engineering (CIFER)(Cat. No. 00TH8520)*, IEEE (pp. 49–57).
- Waltho, C., Elhedhli, S., & Gzara, F. (2019). Green supply chain network design: A review focused on policy adoption and emission quantification. *International Journal of Production Economics*, 208, 305–318. <https://doi.org/10.1016/j.ijpe.2018.12.003>
- Wei, W., Liu, F., & Mei, S. (2015). Energy pricing and dispatch for smart grid retailers under demand response and market price uncertainty. *IEEE Transactions on Smart Grid*, 6, 1364–1374. <https://doi.org/10.1109/TSG.2014.2376522>
- Wind Data Service (2020). Wind Database. <https://www.wind.com.cn/NewSite/edb.html>
- Xiang, D., Qian, Y., Man, Y., & Yang, S. (2014). Techno-economic analysis of the coal-to-olefins process in comparison with the oil-to-olefins process. *Applied Energy*, 113, 639–647.



- Yang, M., & You, F. (2017). Comparative techno-economic and environmental analysis of ethylene and propylene manufacturing from wet shale gas and naphtha. *Industrial & Engineering Chemistry Research*, 56, 4038–4051.
- Ye, W., & You, F. (2016). A computationally efficient simulation-based optimization method with region-wise surrogate modeling for stochastic inventory management of supply chains with general network structures. *Computers & Chemical Engineering*, 87, 164–179. <https://doi.org/10.1016/j.compchemeng.2016.01.015>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.