

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/160040/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Zhao, Yuehao, Li, Zhiyi, Ju, Ping and Zhou, Yue 2023. Two-stage data-driven dispatch for integrated power and natural gas systems by using stochastic model predictive control. *Applied Energy* 343 , 121201. 10.1016/j.apenergy.2023.121201

Publishers page: <http://dx.doi.org/10.1016/j.apenergy.2023.121201>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Two-stage Data-Driven Dispatch for Integrated Power and Natural Gas Systems by Using Stochastic Model Predictive Control

Yuehao Zhao^a, Zhiyi Li^{a,*}, Ping Ju^{a,b}, Yue Zhou^c

^aCollege of Electrical Engineering, Zhejiang University, Hangzhou 310027, China

^bCollege of Energy and Electrical Engineering, Hohai University, Nanjing 211100, China

^cSchool of Engineering, Cardiff University, Cardiff CF24 3AA, U.K

Abstract

The optimal dispatch of the integrated power and natural gas systems can increase the utilization rate of renewable energy and energy efficiency while decreasing operation costs. The common prediction errors of wind power and electric load have the potential to negatively impact the normal operation of the integrated power and natural gas systems. A two-stage data-driven dispatch strategy is proposed to reduce this effect, consisting of the day-ahead dispatch stage and the intraday rolling dispatch stage using stochastic model predictive control (MPC). In the day-ahead dispatch stage, the data-driven chance constraints of tie-line power and reserve of gas-fired generators are built, and the day-ahead tie-line power is obtained and regarded as input parameters to the intraday dispatch stage. In the intraday dispatch stage, the data-driven chance constraints of tie-line power and reserve of gas-fired generators with the latest rolling prediction data are built, and the remaining control variables are obtained. The distribution characteristics of the stochastic prediction errors of wind power and electric load are captured and described by the variational Bayesian Gaussian mixture model with massive historical data. Then the original stochastic mixed-integer nonlinear programming problem is converted to a tractable deterministic one by the quantile-based analytical reformulation and convex relaxation technique. Finally, the proposed strategy is verified by the numerical experiments based on a modified IEEE 33-bus system integrated with a 10-node natural gas system and a micro hydrogen system. The numerical results demonstrate that the proposed strategy reduces the actual costs and decreases the violation rate caused by the stochastic prediction errors of wind power and electric load.

Keywords: two-stage dispatch; chance-constrained programming; data-driven; stochastic model predictive control; integrated power and natural gas systems.

1. Introduction

1.1. Background and motivation

To obtain clean and sustainable energy and address the dilemma of the fossil energy crisis and climate change, the global green energy revolution is taking place all over the world [1]. An increasing proportion of renewable energy such as wind turbines and solar photovoltaic panels is being installed in power systems and integrated power and natural gas systems (IPGS) [2].

The actual output power of renewable energy is closely related to weather conditions. For example, the actual output power of wind turbines relies on wind speed, and the actual output power of photovoltaic panels depends on solar radiation [3,4]. The weather conditions are uncertain and changeable, and thus the actual output power of renewable energy is intermittent, volatile, and stochastic, making it difficult to utilize and dispatch renewable energy effectively. In addition, the actual electric load is also related to the weather conditions, which is stochastic too. As a result, the prediction errors occur as the actual stochastic output of renewable energy and electric load is usually not equal to their prediction values. To make matters worse, the normal operation of integrated power and natural gas systems may be affected by the stochastic prediction errors of wind power and electric load.

The utilization rate of renewable energy and energy efficiency increase and the operation costs decrease with the coordinated and optimal dispatch of IPGS [5]. Therefore, the research on the dispatch of the IPGS is paid increasing attention around the world. IPGS contains numerous emerging multi-

energy conversion devices such as gas-fired generators, power-to-hydrogen (P2H) devices, and various complex multi-energy flows such as electricity and natural gas, as well as stochastic renewable energy. Thus, making the dispatch of IPGS a complex and stochastic optimization problem [6]. As a result, further research is needed on the dispatch for the IPGS considering the stochastic prediction errors of wind power and electric load.

1.2. Literature review

Currently, studies on dispatch for the IPGS are divided into two categories based on whether uncertainty is considered: deterministic dispatch and uncertain dispatch [7].

In deterministic dispatch, all uncertainties from renewable energy and load are ignored, and prediction results are assumed to be completely perfect and accurate, with prediction errors being neglected [6]. However, such dispatch lacks robustness to stochastic wind power [3]. In uncertain dispatch, the day-ahead prediction results of wind power are considered completely accurate, and the prediction errors of renewable energy are neglected. It is assumed that all wind power can be utilized easily, resulting in high adjustment costs and actual operation costs due to the stochastic prediction errors of renewable energy [9].

However, the uncertain and changeable weather conditions make it impossible to obtain completely perfect and accurate prediction data in advance, making prediction errors of renewable energy and load common and inevitable [9,10]. The dispatch results of an IPGS with inaccurate prediction data can be inconsistent with the actual situation, which can easily lead to economic and security problems. As the proportion of renew-

*Corresponding author: Zhiyi Li (zhiyi@zju.edu.cn).

Nomenclature

Abbreviations

CCSP	chance chance-constrainedstic programming
GMM	Gaussian mixture model
IPGS	integrated power and natural gas systems
MISOCP	mix-integer second-order cone programming
MPC	model predictive control
PCC	point of common coupling
P2H	power-to-hydrogen
SVC	static var compensator
VBGMM	variational Bayesian Gaussian mixture model

Sets/Indices

$d \in \Omega_{CB}(j)$	Set of capacitor bank connected to bus j .
$d \in \Omega_{EL}(j)$	Set of electric load connected to bus j .
$g \in \Omega_{GF}(j)$	Set of gas-fired generator connected to bus j .
$g \in \Omega_{GF}(m)$	Set of gas-fired generator connected to gas node m .
$i \in \Omega_{GL}(m)$	Set of gas load connected to gas node m .
$j \in \Omega_{GW}(m)$	Set of gas station connected to gas node m .
$m \in \Omega_{SVC}(j)$	Set of SVC connected to bus j .
$m \in \Omega_{p,in}(m)$	Set of pipeline ingoing gas flow.
$m \in \Omega_{p,out}(m)$	Set of pipeline outgoing gas flow.
$w \in \Omega_{WT}(j)$	Set of wind turbine connected to bus j .
d, q	Index for power demand and the reactive power compensation device.
m, n	Index for gas node m/n .
l, g, w	Index for P2H device, gas-fired generator, and wind turbine.
y, z, t	Index for gas station/ gas demand/ time.

Variables

$C_{op}^{DA}, C_{op}^{ID}, C_{op}^{ID}, C_{op}^{TS}$	Day-ahead/ intraday/the executed operation costs/ two-stage operation costs.
\tilde{e}_t	Prediction errors of wind power and electric load at time t .
$\tilde{e}_{EL,t}$	Prediction errors of electric load at time t .
$\tilde{e}_{w,t}$	Prediction errors of wind power at time t .
f_E, F_E	probability density function /cumulative distribution function
$f_{z,t}^L, f_{g,t}^{GT}$	Gas demand z / gas-fired generator g at time t .
$f_{y,t}^W$	Gas production of the gas station y .
$f_{gm,t}, f_{nm,t}$	Gas flow of the pipeline gm/ mn at time t .
$H_{i,t}$	The generated hydrogen by the P2H i at time t .
$H_{n,t}$	The profit by selling hydrogen.
$f_{y,t}^W$	Gas production of the gas station y .
$f_{gm,t}, f_{nm,t}$	Gas flow of the pipeline gm/ mn at time t .
$I_{ij,t}$	Square of current of branch ij at time t .
$N_{q,t}^{CB}, N_{q,t-1}^{CB}$	The running groups of capacitor bank units at time $t/t-1$.
$P_{d,t}^{EL}, Q_{d,t}^{EL}$	Active/ reactive power of electrical load d at time t .
$P_{g,t}^{GF}, P_{g,t-1}^{GF}$	Output active power of gas-fired generator at time $t/t-1$.
$\tilde{P}_{sub,t}, P_{sub,t}$	Actual/ Scheduled power of tie-line.

$P_{ij,t}, Q_{ij,t}$	Active/ reactive power of branch ij .
$P_{j,t}^N, Q_{j,t}^N$	Injection active/reactive power in node j at time t .
$P_{l,t}^H, P_{l,t-1}^H$	Operating power of P2H device l at time $t/t-1$.
$P_{g,t}^{UR}, P_{g,t}^{DR}$	Upward/ downward reserve capacity of gas-fired generator g .
$Q_{g,t}^{GF}, Q_{g,t-1}^{GF}$	Output reactive power of gas-fired generator g at time $t/t-1$.
$Q_{q,t}^{SVC}, Q_{q,t}^{CB}$	Output reactive power of the SVC/ capacitor bank at time t .
$u_{l,t}^H, u_{l,t-1}^H$	Binary variable that is equal to 1 if P2H device l is ON at time $t/t-1$, and 0 otherwise.
$U_{g,t}^{On}, U_{g,t}^{Off}$	Start-up/ shut-down state of gas-fired generator at time t .
$U_{g,t}^{GF}$	Running state, binary variable that is equal to 1 if gas-fired generator g is ON at time t , and 0 otherwise.
$V_{i,t}, V_{j,t}$	Square of voltage of bus i/ j at time t .
$W_{w,t}^H, W_{w,t}^S$	Consumed wind power by P2H/directly wind power consumed by distribution system at time t .
$\pi_{m,t}, \pi_{n,t}$	Gas pressure in the node m/n at time t .
$\omega_i, \mu_i, \Sigma_i$	Weight/ mean/ variance of component i .
$\Delta S_{q,t}^{CB}$	Switching times of capacitor bank units q at time t .

Parameters

c_{CB}	Cost coefficients of switching of capacitor bank units.
c_y, c_H	The price of natural gas/ hydrogen.
C_{mn}	Constant of the Weymouth equation.
$\bar{f}_{mn}, \underline{f}_{mn}$	Max/min permissible gas flow of the pipeline mn .
$\bar{f}^W, \underline{f}^W$	Max/min permissible gas production of the city gas station.
gap	Relaxation gap.
$\bar{I}_{ij}, \underline{I}_{ij}$	Max/min current square of branch ij .
$\bar{N}^{CB}, N_{max}^{CB}$	The maximum groups/ switching times of capacitor bank units.
N_{max}^{CB}	The number of capacitor bank units.
N_w, N_g, N_E	Total number of wind turbine/ gas-fired generator/ Gaussian components.
$p_{e,t}$	Time-of-use electricity price at time t .
$P_{g,t}^{UR}, P_{g,t}^{DR}, P_{g,t}^S$	Cost coefficients of upward/downward reserve/ startup of gas-fired generator g .
\bar{P}_g^{GF}, P_g^{GF}	Max/min output power of gas-fired generator g .
$P_{g,t}^{GF}, P_{g,t-1}^{GF}$	Max upward/downward ramping capacity of gas-fired generator g at time t .
$\tilde{P}_t^{EL}, P_t^{EL}$	Actual/Predicted output of the load.
$P_{l,max}^H, P_{l,min}^H$	Max/min operating power of P2H device l .
P_{sub}^{max}	The maximum permissible value of the tie-line power.
q_q^{CB}	The capacity of capacitor bank q .
$\bar{Q}_q^{SVC}, Q_q^{SVC}$	Max/min output power of SVC q .
Q_U, Pr	Quantile/ Probability
\bar{V}, \underline{V}	Max/min voltage square of the bus.
r_{ij}, x_{ij}	Resistance/ reactance of branch ij .
τ	Dispatch period, 15 minutes*96.
$\tilde{W}_{w,t}^a, W_{w,t}^f$	Actual/Predicted output of the wind turbine w .
$\alpha_{sub}^{max}, \alpha_{UR}, \alpha_{DR}$	The tolerance level for overload/ insufficient upward/downward reserve.
$\bar{\pi}, \underline{\pi}$	Max/min gas pressure of the node.
$\eta_h^P, \eta_h^S, \eta_{g,t}^{GF}$	The conversion efficiency of PEMEC/ SOEC, or gas-fired generator g at time.

Other notations are defined in the text.

-able energy and thermostatically controlled loads increases, the uncertainty of the IPGS also grows, making uncertain dispatch even more important.

There is an acknowledged fact, that the prediction errors of renewable energy and electric load usually increase as the forecast time horizon prolongs [12]. A shorter forecast time horizon often means more accurate prediction data and lower

prediction errors [13]. As a result, the model predictive control (MPC) or rolling optimization method in a receding-horizon manner is gradually and widely applied to the dispatch of the IGPS to handle its uncertainty [12]. The latest prediction data of renewable energy and load is regarded as input data of the MPC method. Compared to the day-ahead deterministic dispatch, the intraday dispatch results based on MPC are closer to the actual situation. In Ref. [12], the rolling MPC method is used to handle the operational uncertainties of the distributed renewable energy and multi-energy load, and the re-scheduling cost is lower due to more accurate prediction data. However, the actual output of the wind power and solar power, and the load value are regarded as deterministic, and the rolling updated ultra-short-term forecasting data is input into the energy management system to handle the uncertainty [12]. The rolling-horizon approach is used to alleviate the power deficit because of the wind power, but its operation and adjustment costs are still large [9].

The MPC possesses a certain degree of robustness to the uncertainties of the renewable energy and load, but it is a deterministic dispatch in essence with the rolling latest prediction data [14]. This reference clearly stated that the deterministic formulation of MPC typically renders it inherently inadequate for systematically dealing with uncertainties. Specifically, The inherent deterministic characteristics of the MPC method limit its robustness to uncertainties. As a result, it cannot adapt to random changes in renewable energy and load in IPGS [14]. To address the above problems, the dispatch based on the stochastic MPC is proposed recently. In Ref. [15], the stochastic MPC is adopted to coordinate the optimal operation of a multi-energy microgrid, which can decrease the operation costs and guarantee the system security of the multi-energy microgrid. The stochastic prediction errors of wind turbines and photovoltaic panels are assumed to follow the classical Bata distribution, and the load is assumed to follow the Gaussian distribution. The stochastic scenario-based MPC dispatch is used to deal with uncertainties. In Ref. [16], the prediction errors of the wind turbines, photovoltaic panels, and load are assumed to follow the Gaussian distribution, and the stochastic MPC is used to handle the stochastic fluctuation of the voltage.

Many previous studies on the dispatch of IPGS using traditional stochastic MPC have made a strong assumption about the stochastic prediction errors distribution of renewable energy and electric load, that they follow a common and relatively simple distribution such as Gaussian or Beta distribution [15,16]. It is a fact that the prediction errors in the different geographical regions may have unique characteristics, which cannot be described by the same and simple distribution. For instance, the stochastic prediction errors of renewable energy and electric load usually show multi-peak and asymmetry characteristics, but the Gaussian distribution is unimodal and symmetrical [17]. The dispatch results based on the inaccurate prediction data easily lead to uneconomical and unreliable operation of IPGS.

Besides, those strong common parametric probabilistic assumption such as Gaussian distribution in the above previous stochastic MPC is only applicable to special scenarios, but it is

not suitable for IPGS the stochastic renewable energy and load [18]. The common parametric probabilistic distribution has strong priori hypotheses, that its parameters are fixed and very limited. Numerous potential information of the historical prediction data and measured data has not been fully utilized, causing serious waste of the data resource of renewable energy and load. To better utilize those data and describe their distribution characteristic, the Gaussian mixture model (GMM) is adopted recently [11,12]. The GMM is a data-driven fitting method, which can utilize the historical data of the prediction errors and describe its distribution characteristic accurately. However, the describing or fitting effect of the GMM method relies largely on the appropriate number of the Gaussian component. The number of the Gaussian component is often selected by manual observation and trial, which can lead to overfitting or underfitting phenomena [19]. As a result, how to better describe the distribution characteristics of the stochastic prediction errors of renewable energy and electric load with full utilization of those data and integrate it into the stochastic MPC of the IPGS needs further research.

1.3. Contributions

A two-stage data-driven dispatch for integrated power and natural gas systems using stochastic MPC is proposed in this paper. The main contributions are summarized as follows:

1) A two-stage dispatch strategy of IPGS is proposed, consisting of the day-ahead dispatch and intraday rolling dispatch by using the stochastic MPC. The stochastic MPC with the latest prediction data is used to deal with the uncertainty of wind power and electric load in IPGS effectively.

2) A data-driven chance-constrained dispatch strategy is proposed to further mitigate adverse effects caused by the stochastic actual output and prediction errors of wind power and electric load. The variational Bayesian Gaussian mixture model (VBGM) is adopted to accurately describe the distribution characteristic of the prediction errors of wind power and electric load [20]. Additionally, the tie-line power and the reserve capacity of gas-fired generators are modeled as chance constraints.

3) A comprehensive evaluation index of the violation rate is proposed to assess the effectiveness of the proposed dispatch strategy. This index includes both the maximum violation rate and the cumulative value of the violation rate, which provides a more reasonable and intuitive measure of the violation probability.

The rest of this paper is organized as follows. Section 2 presents the two-stage dispatch strategy framework using stochastic MPC. The stochastic optimization of IPGS is described in Section 3. Section 4 outlines the solution method. Finally, conclusions are presented in Section 5.

2. Two-stage dispatch strategy of IPGS

This section starts by introducing the schematic of the IPGS. Next, a two-stage dispatch strategy for the IPGS is proposed, which includes both day-ahead and intraday dispatches by using the stochastic MPC. For the convenience of writing, “the

prediction errors wind power and electric load” is abbreviated as “prediction errors” in the following section.

2.1. The schematic of the IPGS

In this work, the structure of the IPGS is schematically illustrated in Fig. 1. The IPGS is connected to the external grid through the point of common coupling (PCC) by a tie-line and connected to a gas station. It mainly consists of two subsystems, i.e., a power distribution system that integrates the P2H devices and a natural gas system.

In many countries such as the U.K., the power and natural gas systems are operated and managed jointly by a single utility [21]. Therefore, we assume that the multi-energy operator is responsible for the energy management of the IPGS and dispatches the various energy conversion devices to meet multi-energy demands. The power demand is met by the output power of the wind turbine and gas-fired generator, as well as the purchased power from the external power grid. The P2H devices consume wind power to generate hydrogen, which is then sold to external hydrogen buyers for profit. The gas-fired generator consumes natural gas, and the natural gas demand is met by the gas station.

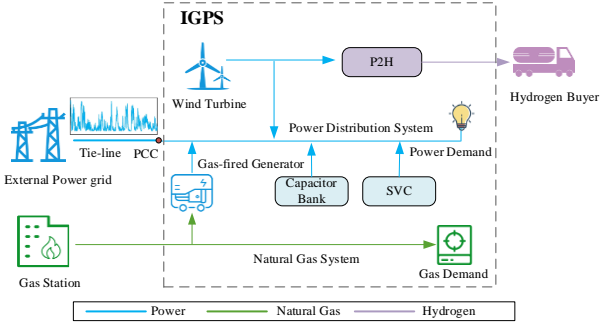


Fig. 1 Structure of IPGS

2.2. Two-stage dispatch strategy

In IGPS, the actual output of wind power is stochastic due to the changeable weather conditions, and the electric load is also stochastic. To address these uncertainties and mitigate the adverse effects of prediction errors, a two-stage dispatch strategy is proposed. The framework of the two-stage dispatch is shown in Fig. 2.

The first stage, which we refer as day-ahead dispatch, determines the tie-line power and the number of capacitor bank units in advance [22,23]. The multi-energy operator needs to decide the day-ahead tie-line power in advance, and provide this value to the external power grid, which is beneficial for the external power grid to prepare the tie-line power and the economical operation of the external power grid. In return, the multi-energy operator can get a favorable day-ahead electricity price from the external power grid. Besides, the number of capacitor bank units in operation is usually decided as we assume the effect of prediction errors on the reactive power is ignored. The day-ahead dispatch is divided into 96 equal time (15-min interval), and the tie-line power and number of capacitor bank units will be adopted in the next stage.

The second stage, which we refer as to intraday rolling

dispatch by using stochastic MPC, determines the remaining control variables. The prediction of wind power generation and electric load in this stage is usually more accurate, and many devices are flexible to adjust their running state, so numerous remaining control variables are usually decided in this stage. The intraday rolling dispatch is divided into 96 equal time (15-min interval). The rolling horizon T_{ID} is set as 4 hours (15 mins*4*4) according to the latest prediction data four hours in advance, and the intraday dispatch will roll repeatedly 24 times (once an hour) in a day. In other words, the intraday rolling dispatch repeats every hour. However, considering the accuracy of the prediction data of the wind power and electric load, only the decision variables in the first hour (15 mins*4) in the rolling horizon will be adopted and executed. This procedure is presented in Fig. 2, the control variables in the first hour of the rolling horizon, denoted by the circle with bright green color will be executed. Besides, the remaining control variables in the last three hours of the rolling horizon, denoted by the circle with grey color will not be executed.

2.2.1. Day-ahead dispatch objective

The day-ahead dispatch strategy is formulated according to the day-ahead prediction data of the wind power and load. The IGPS needs to purchase power from the external grid through the tie-line. To mitigate the adverse effects of fluctuating tie-line power on the external grid, the power purchased from the external grid considers the uncertainty of the stochastic prediction errors, and the data-driven chance-constrained stochastic programming (CCSP) is adopted to handle the stochastic prediction errors which will be introduced in detail in the next Section 3.

The objective function of the day-ahead dispatch is to minimize the operation costs of the multi-energy operator for the next day, which consists of the power and natural gas purchase costs, start-up and reserve capacity costs of the gas-fired generator, profits from generating hydrogen, and switching costs of the capacitor bank units. This function is represented in Equation (1).

$$\min C_{op}^{DA} = \sum_{t=1}^T \left(\sum_{g=1}^{N_g} (p_g^{UR} P_{g,t}^{UR} + p_g^{DR} P_{g,t}^{DR} + p_g^S P_{g,t}^{On}) + p_{e,t} P_{sub,t} \right) + \sum_{y=1}^{N_w} c_y f_{y,t}^w + \sum_{s=1}^{N_{CB}} c_{CB} \Delta S_{CB,s} - c_H H_{B,t} \quad (1)$$

2.2.2. Intraday rolling dispatch objective using stochastic MPC

The intraday rolling dispatch strategy by using stochastic MPC is formulated according to the latest intraday rolling prediction data of wind power and electric load. Compared to the day-ahead prediction data, the accuracy of the latest intraday rolling prediction data of wind power and electric load is higher, so the remaining control variables are decided in this stage, which mainly refers to start-up and reserve capacity costs of the gas-fired generator, natural gas purchase costs, profits from generating hydrogen,

$$\min C_{op}^{ID} = \sum_{t=1}^{T_{ID}} \left(\sum_{g=1}^{N_g} (p_g^{UR} P_{g,t}^{UR} + p_g^{DR} P_{g,t}^{DR} + p_g^S P_{g,t}^{On}) + \sum_{y=1}^{N_w} c_y f_{y,t}^w - c_H H_{B,t} \right) \quad (2)$$

The objective function of intraday rolling dispatch based on stochastic MPC is minimizing the operating costs of the multi-energy operator in the rolling horizon, which consist of power and natural gas purchase costs, the start-up and reserve capacity costs of the gas-fired generator, and profits by generating hydrogen, as presented by (2).

2.2.4. Total budgeted operation costs of the two-stage dispatch

After obtaining the decision variables through optimization and calculation of day-ahead dispatch and intraday rolling

dispatch, the total budgeted operation costs of the multi-energy operator in the two-stage dispatch can be calculated using equation (3). The budgeted operation costs consist of two parts: the switching costs of the capacitor bank units in day-ahead dispatch and the total executed operation costs in intraday rolling dispatch over 24 hours. It should be noted that the term "budgeted operation costs" refers to the costs estimated based on the day-ahead and intraday rolling prediction data of wind power and electric load, rather than the minimized objective functions like (1) or (2).

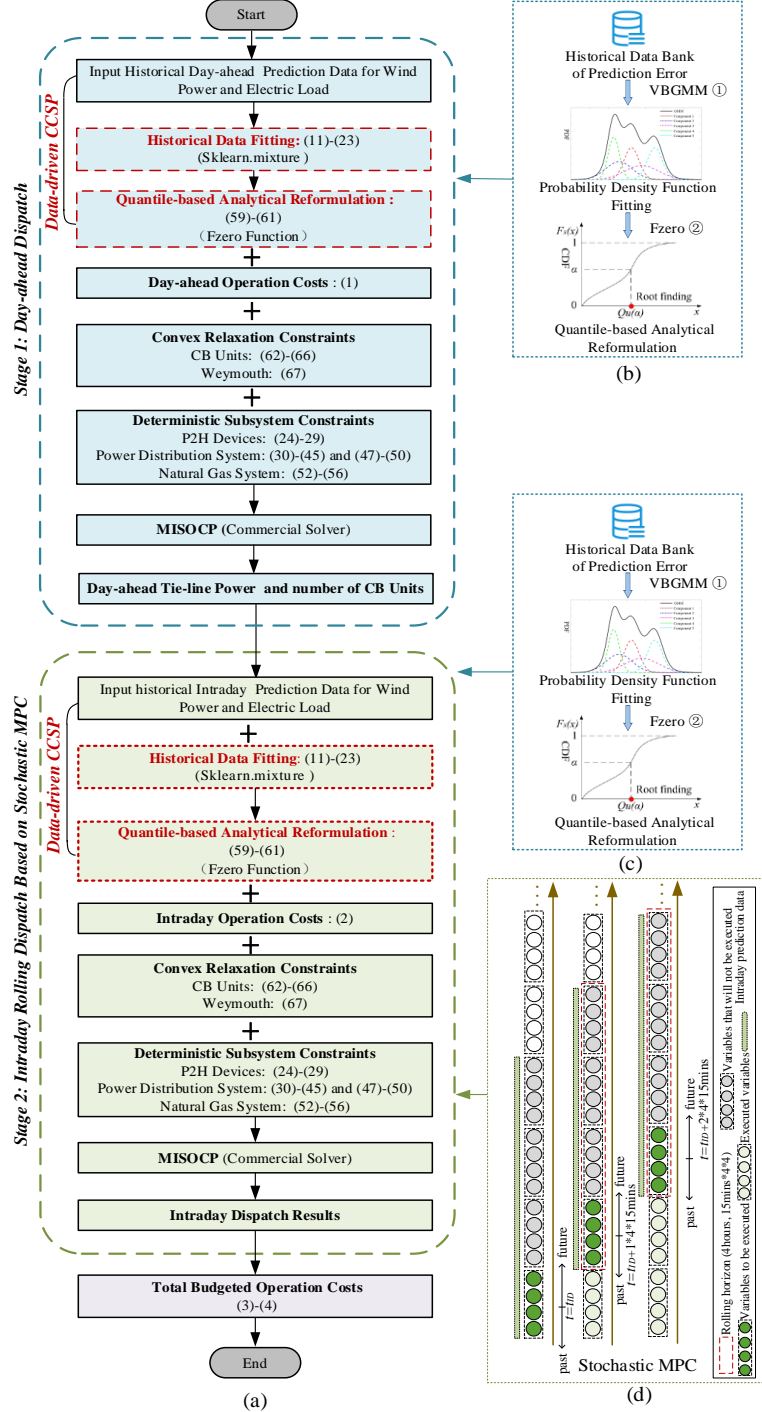


Fig. 2 Two-stage dispatch framework

$$C_{op}^{TS} = \sum_{t=1}^T (p_{e,t} P_{sub,t} + \sum_{q=1}^{N_{CB}} c_{CB} \Delta S_{q,t}^{CB}) + \sum_{t=1}^{24} C_{op,t}^{IDE} \quad (3)$$

where the executed operation costs at hour t are equal to the operation in the first hour (15 mins*4) of the rolling horizon, which are calculated by (4).

$$C_{op,t}^{IDE} = \sum_{t=1}^4 \left(\sum_{g=1}^{N_g} (p_g^{UR} P_{g,t}^{UR} + p_g^{DR} P_{g,t}^{DR} + p_g^S P_{g,t}^{On}) + \sum_{y=1}^{N_w} c_y f_{y,t}^W - c_H H_{B,t} \right) \quad (4)$$

3. The stochastic dispatch strategy

3.1. Data-driven chance constraints

With an increasing proportion of wind power and thermostatically controlled loads in IGPS, the stochastic output of the wind power is easy to lead to the fluctuation of the tie-line power, which affects the economic operation of the external power grid. To mitigate the adverse effects of the fluctuation of stochastic wind power and electric load, the data-driven CCSP is built, which consists of two parts, stochastic chance constraints and data-driven distribution characteristic fitting method of prediction errors. In the above two-stage stochastic dispatch strategy, the data-driven CCSP is one of the vital constraints to handle stochastic prediction errors.

3.1.1. Chance constraints of tie-line and reserve capacity

To reduce the adverse effect of stochastic tie-line power on the economic operation of the external power grid, the fluctuation of the tie-line needs to be limited [24]. Without loss of generality, assume the gas load is deterministic, and the influence of stochastic prediction errors on the change of network losses is negligible. The stochastic prediction errors consist of two parts, the prediction error of wind power and electric load. The actual output of the wind power equals the predicted wind power plus its stochastic prediction errors, which is presented in (5). Similarly, the actual value of the electric load is equal to the predicted value plus stochastic prediction errors, as shown in equation (6). The stochastic tie-line power, which is affected by the stochastic prediction errors, is modeled as chance constraints and is presented in equations (7) and (8).

$$\tilde{W}_{w,t}^a = W_{w,t}^f + \tilde{e}_{w,t} \quad (5)$$

$$\tilde{P}_t^{EL} = P_t^{EL} + \tilde{e}_{EL,t} \quad (6)$$

$$\tilde{P}_{sub,t} = P_{sub,t} - \tilde{e}_t \quad (7)$$

$$\Pr(\tilde{P}_{sub,t} \leq P_{sub}^{\max}) \geq 1 - \alpha_{sub}^{\max} \quad (8)$$

In addition, due to the stochastic output of the wind power and electric load, a reasonable reserve capacity of gas-fired generator is required for the power balance. The chance constraints of upward/downward reserve capacity of the gas-fired generator are presented as (9) and (10) respectively.

$$\Pr\left(\sum_{g=1}^{N_g} P_{g,t}^{UR} \geq \sum_{w=1}^{N_w} W_{w,t}^s - \sum_{w=1}^{N_w} (\tilde{W}_{w,t}^a - W_{w,t}^H) + \tilde{e}_{EL,t}\right) \geq 1 - \alpha_{UR} \quad (9)$$

$$\Pr\left(\sum_{g=1}^{N_g} P_{g,t}^{DR} \geq -\sum_{w=1}^{N_w} W_{w,t}^s + \sum_{w=1}^{N_w} (\tilde{W}_{w,t}^a - W_{w,t}^H) - \tilde{e}_{EL,t}\right) \geq 1 - \alpha_{DR} \quad (10)$$

3.1.2. Data-driven fitting method for the prediction errors

As many advanced measurement devices are widely installed in the IGPS, there are numerous historical prediction data and measured data, but now its potential key statistic information has not been well mined and utilized. For example, masses of historical prediction data and measured data of the wind power and electric load have been accumulated, but parametric Gaussian distribution is still assumed as the exact distribution of stochastic prediction errors in many references, causing serious waste of data information resources [15,16]. To make matters worse, the inaccurate distribution characteristic of the stochastic prediction errors easily hurts the operational economy and safety of the IGPS [10,11].

The distribution characteristic of the stochastic prediction errors is accurately described and fitted by the VBGMM in this work. In theory, an arbitrary probability distribution can be fitted relatively accurately by the GMM if it adjusts its parameters, such as the number of components, weight, means, and covariance [10]. However, as it is hard to select a reasonable component number and parameters, the overfitting or underfitting problem is inevitable. Compared to the traditional Gaussian distribution and GMM method, the fitting effect of the VBGMM method is usually better as it can select the component number and parameters automatically. The VBGMM is a data-driven fitting method that utilizes the historical data to fit the distribution characteristic, and the utilization ratio of the historical data is improved. The VBGMM is a nonparametric Bayesian model based on the Dirichlet process, which can infer the number of Gaussian components of the prediction errors according to its historical data [25]. Besides, some remaining vital parameters are also inferred by the VBGMM such as the weight, mean, and variance. Once these parameters are obtained, the probability density function and cumulative distribution function of the stochastic prediction errors are also obtained, which are then introduced in the CCSP. The flowchart of the VBGMM method is presented in Fig. 3.

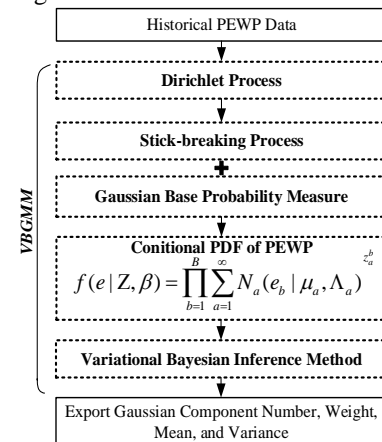


Fig. 3 Flowchart of the VBGMM method.

For finite partitions $\theta_1, \dots, \theta_i$ in measure space Θ , if the distribution G obeys the rule (11), then G follows the Dirichlet process, which is presented by (12) [26]. The Dirichlet process G consists of two key elements, i.e., concentration parameter ϕ and the base probability measure G_0 . The base probability

measure in VBGMM is the Gaussian distribution.

$$G((\theta_1), \dots, (\theta_i)) \sim \text{Dir}(\phi G_0(\theta_1), \dots, \phi G_0(\theta_i)) \quad (11)$$

$$G \sim \text{DP}(\phi, G_0) \quad (12)$$

The stick-breaking process is one of the common methods to construct the Dirichlet process, and it is adopted in this work. The stick-breaking process is built in the following (13)-(16). The random sampling of the base probability measure is used to calculate the categorization weight of the component, and the computing formula of Dirichlet process G is shown (16). The hidden variables in VBGMM are presented in (17).

$$v_a \sim \text{Beta}(v_a | \mathbf{1}, \phi) \quad (13)$$

$$\omega_a = v_a \prod_{s=1}^{k-1} (1 - v_s) \quad (14)$$

$$\sum_{a=1}^{\infty} \omega_a = 1 \quad (15)$$

$$G = \sum_{a=1}^{\infty} \omega_a \delta_{\theta_a} \quad (16)$$

$$\mathfrak{V} = \{v_a\}_{a=1}^{\infty}, \mathfrak{Y} = \{\omega_a\}_{a=1}^{\infty} \quad (17)$$

The accumulated historical data of the prediction errors $E = \{e_b\}_{b=1}^B$ is considered a set of stochastic variables. Then the joint probability density function of prediction errors is obtained as denoted in (18) [16].

$$f(e | \omega, \beta) = \prod_{b=1}^B \sum_{a=1}^{\infty} \omega_a N_a(e_b | \mu_a, \Lambda_a) \quad (18)$$

where $\omega = \{\omega_a\}_{a=1}^{\infty}$ and $\beta = \{\mu_a, \Lambda_a\}_{a=1}^{\infty}$.

The categories to which stochastic variables in the historical data of the prediction errors are classified by the binary indicator variable z_a^b . The conditional probability density function with indicator variable is denoted in (19) [19].

$$f(e | Z, \beta) = \prod_{b=1}^B \sum_{a=1}^{\infty} N_a(e_b | \mu_a, \Lambda_a)^{z_a^b} \quad (19)$$

where $Z = \{z_a^b\}_{a=1, b=1}^{a=\infty, b=B}$.

Then the posterior distribution parameter of the hidden variable $\Xi = \{Z, \omega, \mathfrak{V}, \mathfrak{Y}\}$ is estimated by the variational Bayesian inference method [19]. The variational Bayesian inference method is implemented and coded based on sklearn.mixture package [20]. The inferred cumulative distribution function and probability density function of the prediction errors are presented in (20) and (21) respectively. Besides, the sum of the weights of all Gaussian components is 1.

$$f_E(\tilde{e}_i) = \sum_{e=1}^{N_E} \omega_e N_e(\tilde{e}_i | \mu_e, \Sigma_e) \quad (20)$$

$$F_E(\tilde{e}_i) = \sum_{i=1}^{N_E} \omega_i \Phi_i(\tilde{e}_i | \mu_i, \Sigma_i) \quad (21)$$

$$\tilde{e}_i = [\tilde{e}_{w,i}, -\tilde{e}_{EL,i}] \quad (22)$$

$$\sum_{i=1}^{N_E} \omega_i = 1, \omega_i \geq 0 \quad (23)$$

3.2. Operation constraints of P2H devices

1) Ramp constraints of P2H

$$P_{l,t}^H - P_{l,t-1}^H \leq P_{l,\text{up}}^H \quad (24)$$

$$P_{l,t-1}^H - P_{l,t}^H \leq P_{l,\text{down}}^H \quad (25)$$

where (24) and (25) denote the ramp up and ramp down constraints of the P2H device.

2) Operation power constraints of P2H

$$u_{l,t}^H P_{l,\text{min}}^H \leq P_{l,t}^H \leq u_{l,t}^H P_{l,\text{max}}^H \quad (26)$$

$$U_{l,t}^{\text{On}} - U_{l,t}^{\text{Off}} = u_{l,t-1}^H - u_{l,t}^H \quad (27)$$

where (26)-(27) denote the operating power constraints of P2H device, respectively.

5) Conversion efficiency of P2H

$$H_{l,t} = \eta_l^H P_{l,t}^H \quad (28)$$

where (28) denotes the conversion efficiency of the P2H device, respectively.

6) Hydrogen balance constraints

$$\sum_{l \in \Omega_H(l)} H_{l,t} = H_{B,t} \quad (29)$$

where (29) denotes generated hydrogen that is provided to the external hydrogen buyer.

3.3. Power distribution system constraints

1) Wind power allocation constraints

$$W_{w,t}^s = W_{w,t}^f - W_{w,t}^H \quad (30)$$

$$W_{w,t}^H = \sum_{l \in \Omega_H(l)} P_{l,t}^H \quad (31)$$

where (30) denotes the relationship between the wind power directly scheduled and consumed by the power distribution systems, wind power consumed by the P2H devices, and the predicted value of wind power. The total wind power consumed by the P2H device is described in (31).

2) Branch power flow constraints

$$P_{ij,t} - r_{ij} I_{ij,t} + P_{j,t}^N = \sum_{k \in c(j)} P_{jk,t} \quad (32)$$

$$Q_{ij,t} - x_{ij} I_{ij,t} + Q_{j,t}^N = \sum_{k \in c(j)} Q_{jk,t} \quad (33)$$

$$V_{j,t} = V_{i,t} - 2(P_{ij,t} r_{ij} + Q_{ij,t} x_{ij}) + (r_{ij,t}^2 + x_{ij,t}^2) I_{ij,t} \quad (34)$$

$$V_{i,t} I_{ij,t} \geq P_{ij,t}^2 + Q_{ij,t}^2 \quad (35)$$

where (32)-(35) denote the branch power flow.

3) Nodal power balance constraints

$$P_{j,t}^N = \sum_{g \in \Omega_{GT}(j)} P_{g,t}^{\text{GF}} + P_{\text{sub},t} + \sum_{w \in \Omega_{WT}(j)} W_{w,t}^s - \sum_{d \in \Omega_{EL}(j)} P_{d,t}^{\text{EL}} \quad (36)$$

$$Q_{j,t}^N = \sum_{g \in \Omega_{GT}(j)} Q_{g,t}^{\text{GT}} + Q_{\text{sub},t} + \sum_{q \in \Omega_{SVC}(j)} Q_{q,t}^{\text{SVC}} + \sum_{q \in \Omega_{CB}(j)} Q_{q,t}^{\text{CB}} - \sum_{d \in \Omega_{EL}(j)} Q_{d,t}^{\text{EL}} \quad (37)$$

where (36) and (37) denote the nodal active and reactive power balance constraints, respectively. The nodal active power equals the sum of the output of gas-fired generator and wind power, substation (if any) minus the load; The nodal reactive power equals the sum of the output of gas-fired generator, SVC, capacitor bank, and substation (if any) minus the load.

4) Operation constraints of the gas-fired generator

$$U_{g,t}^{\text{On}} - U_{g,t}^{\text{Off}} = U_{g,t}^{\text{GF}} - U_{g,t-1}^{\text{GF}} \quad (41)$$

$$0 \leq U_{g,t}^{\text{On}} \leq U_{g,t}^{\text{GF}} \quad (42)$$

$$0 \leq U_{g,t-1}^{\text{Off}} \leq 1 - U_{g,t}^{\text{GF}} \quad (42)$$

where (38)-(40) denote the relations among start-up, shut-down,

and running indicators of gas-fired generator in two consecutive time intervals. And the ramp up/down constraints of gas-fired generator are presented as follows:

$$P_{g,t}^{GF} - P_{g,t-1}^{GF} \leq P_{g,up}^{GF} \quad (41)$$

$$P_{g,t-1}^{GF} - P_{g,t}^{GF} \leq P_{g,down}^{GF} \quad (42)$$

The operating power constraints with the reserve of gas-fired generator are stated as follows:

$$U_{g,t}^{GF} P_{g,t}^{GF} \leq P_{g,t}^{GF} - P_{g,t}^{DR} \quad (43)$$

$$U_{g,t}^{GF} P_{g,t}^{GF} \leq P_{g,t}^{GF} - P_{g,t}^{DR} \quad (44)$$

where (43) and (44) denote the upward/downward reserve.

5) Operation constraints of capacitor bank units[27]

$$Q_{q,t}^{CB} = N_{q,t}^{CB} q_q^{CB} \quad (45)$$

$$\sum_{t=1}^T |N_{q,t}^{CB} - N_{q,t-1}^{CB}| \leq N_{max}^{CB} \quad (46)$$

$$0 \leq N_{q,t}^{CB} \leq \bar{N}^{CB} \quad (47)$$

where (45) describes the nodal total reactive power generated by capacitor bank. (46) states the maximum total switching times of capacitor bank in the entire dispatch period. (47) state the maximum number of running groups of capacitor bank units.

6) Operation constraints of SVC

$$\underline{Q}_q^{SVC} \leq Q_{q,t}^{SVC} \leq \bar{Q}_q^{SVC} \quad (48)$$

where (48) describes the operating power range of SVC.

7) Operation constraints of the network

$$I_{ij} \leq I_{ij,t} \leq \bar{I}_{ij} \quad (49)$$

$$\underline{V} \leq V_{i,t} \leq \bar{V} \quad (50)$$

where (49) and (50) constrain the operating range of the branch current of the line and bus voltage.

3.4. Natural gas system constraints

1) Pipeline natural gas flow constraints

The relationship between the flow of pipeline and the gas pressure of node is often denoted by the Weymouth equation, as shown in (51) [28].

$$f_{mn,t}^2 = C_{mn}^2 (\pi_{m,t}^2 - \pi_{n,t}^2) \quad (51)$$

2) Nodal natural gas flow balance constraints

$$\sum_{m \in \Omega_{p,(m)}} f_{gm,t} - \sum_{m \in \Omega_{p,(m)}} f_{mn,t} + \sum_{y \in \Omega_{GW(m)}} f_{y,t}^W = \sum_{z \in \Omega_{GL(m)}} f_{z,t}^L + \sum_{g \in \Omega_{GT(m)}} f_{g,t}^{GT} \quad (52)$$

where (52) represents the gas flow balance at each node.

3) Operation constraints

$$\underline{\pi} \leq \pi_t \leq \bar{\pi} \quad (53)$$

$$\underline{f}_{mn} \leq f_{mn,t} \leq \bar{f}_{mn} \quad (54)$$

$$\underline{f}^W \leq f_{y,t}^W \leq \bar{f}^W \quad (55)$$

where (53) and (54) constrain the operating range of node gas pressure and pipeline gas flow. where (55) denotes the operating range of gas station.

4) Natural gas consumption of the gas-fired generator [28]

The natural gas consumed by the gas-fired generator is

usually described by a linear conversion model, which is denoted by the following (56).

$$f_{g,t}^{GF} = P_{g,t}^{GF} \eta_{g,t}^{GF} \quad (56)$$

4. Solution Method

The day-ahead dispatch strategy is summarized as (57), and the intraday dispatch strategy is summarized as (58).

$$\begin{cases} \min C_{op}^{DA} \\ \text{s.t. (5)-(10), (20)-(56)} \end{cases} \quad (57)$$

$$\begin{cases} \min C_{op}^{ID} \\ \text{s.t. (5)-(10), (20)-(56)} \end{cases} \quad (58)$$

where (5)-(10) denote chance constraints, (20)-(23) are the fitted probability density function and cumulative distribution function, (24)-(29) describe the operation constraints of P2H devices, (30)-(50) state the operation constraints of power distribution system, and (51)-(56) are natural gas system constraints.

As massive prediction and measured data of renewable energy and electric load accumulated in IPGS, its prediction errors can be obtained easily according to (5) and (6), and these prediction errors are also stored in the historical data bank of IPGS. Then its distribution characteristics are inferred by VBGM, as shown in the first step in (b) and (c) of Fig. 2.

4.1. Quantile-based analytical reformulation

The chance constraints in (8)-(10) are nonconvex, which cannot be solved by the common commercial solvers. The stochastic $\tilde{w}_{w,t}^a$ in (9)-(10) are substituted by the (5) and (30), respectively. The quantile-based analytical reformulation is adopted, then the nonconvex chance constraints are converted to the tractable deterministic constraints, as shown in (59)-(61).

$$P_{sub,t} \leq P_{sub}^{max} + Q_U(\alpha_{sub}^{max} | \tilde{e}_t) \quad (59)$$

$$\sum_{g=1}^{N_G} P_{g,t}^{UR} \geq -Q_U(\alpha_{UR} | \tilde{e}_t) \quad (60)$$

$$\sum_{g=1}^{N_G} P_{g,t}^{DR} \geq Q_U(1 - \alpha_{DR} | \tilde{e}_t) \quad (61) \text{ where (59)-(61) are}$$

the tractable deterministic constraints of tie-line power, upward reserve capacity, and downward reserve capacity.

The cumulative distribution function of the stochastic prediction errors in (21) can be regarded as a nonlinear equation. The quantile value of the stochastic prediction errors is the root of the cumulative distribution function in the corresponding confidence level α , and many root-finding methods for the nonlinear equation can be used to find the root such as the Newton method [10]. As the fzero function can find the root fast and accurately, it is adopted to find the root of cumulative distribution function (quantile value) [29], as shown in the second step in (b) and (c) of Fig. 2.

4.2. Convex relaxation technique

The absolute value in (56) is nonconvex which can be handled by the linear equivalent substitute with auxiliary variables $N_{q,t}^+$ and $N_{q,t-1}^-$, as shown in (62)-(66) [27].

$$\sum_{t=1}^T (N_{q,t}^+ + N_{q,t-1}^-) \leq N_{\max}^{\text{CB}} \quad (62)$$

$$\sum_{t=1}^T (N_{q,t}^+ + N_{q,t-1}^-) \leq N_{\max}^{\text{CB}} \quad (63)$$

$$N_{q,t}^{\text{CB}} - N_{q,t-1}^{\text{CB}} = N_{q,t}^+ - N_{q,t-1}^- \quad (64)$$

$$N_{q,t}^+ \geq 0, N_{q,t}^- \geq 0 \quad (65)$$

$$\Delta S_{q,t}^{\text{CB}} = N_{q,t}^{\text{CB}} - N_{q,t-1}^{\text{CB}} \quad (66)$$

The nonlinear Weymouth equation (62) is converted into the following cone constraints (67) by the second-order cone programming technique. However, this cone relaxation technique is not our contribution and focus, and more detailed descriptions are discussed in [30].

$$f_{mn,ave,t}^2 + C_{mn}^2 \pi_{n,t}^2 \leq C_{mn}^2 \pi_{m,t}^2 \quad (67)$$

Then the original mixed-integer nonlinear programming problem is converted into a tractable deterministic mix-integer second-order cone programming (MISOCP) problem based on the above quantile-based analytical reformulation and convex relaxation technique. It is easy for common commercial solvers to solve this MISOCP problem. The flowchart of the solution method is presented in Fig. 2. After the two-stage optimization dispatch, the values of all decision variables are obtained, and the IGPS is in optimal operation state, and its total budgeted operation costs are calculated by (3) and (4).

4.3. Solution error analysis

The potential solution errors or gap may occur because of the above quantile-based analytical reformulation and the convex relaxation technique. The default termination tolerance of the fzezo function is $\text{eps}, 2.2204\text{e}-16$, so its calculation gap can be ignored [29]. As a result, the potential error of quantile-based analytical reformulation is also neglected.

As for the second one, convex relaxation, consists of two parts, absolute value and second-order cone relaxation. The relaxed linear equations with auxiliary variables are completely equivalent to their original ones, and the relaxation gap is zero. As for the potential errors due to the cone relaxation technique, and the following equation is used to calculate the maximum relaxation gap of the pipeline mn [31]. The solution in the relaxed MISOCP model can be regarded as exact if the relaxation gap obtained from the following equation is small enough.

$$\text{gap} = \max \left\{ \frac{C_{mn}^2 \pi_{m,t}^2 - C_{mn}^2 \pi_{n,t}^2 - f_{mn,t}^2}{f_{mn,t}^2}, \forall t \in T, \forall mn \right\} \quad (68)$$

where gap is the maximum relaxation gap.

5. Case Study

5.1. Case description

The programs of the proposed two-stage dispatch strategy are written in Julia/ JuMP [32] environment and executed on a laptop with AMD R7-4800U (1.8 GHz) processors and 16 GB RAM. The day-ahead dispatch and intraday rolling dispatch are solved by the commercial solver Gurobi (ver. 9.0.3) in which the relative optimality gap is set at 0.

The proposed strategy is tested on the IPGS shown in Fig. 4,

which is based on the modified IEEE 33-bus system [33] integrated with a 10-node natural gas system [34] and a micro hydrogen system. As for the power distribution system, the base value of power is 1 MVA, and voltage deviation limits are set to be $\pm 5\%$ of the rated level (4.16 kV). Bus 1 represents the substation, connecting to the external power grid through a tie-line. Bus 31 installs one wind farm, one PEMEC, and two SOECs. Capacitor bank units are installed at bus 19 and bus 23, respectively. Besides, the SVC is located at bus 16 and bus 33. There are two gas-fired generators located at bus 3 and bus 6. The 10-node natural gas system includes one natural gas station in node 1, and node 2 and node 6 are responsible for supplying natural gas to gas-fired generators. The maximum permissible value of the active power of the tie-line is set as 6.8 MW. The tolerance level for stochastic risks in chance constraints is set as 5%. The historical data of the actual prediction data and measured data of wind power are taken from the Belgian's electricity system operator, Elia [35].

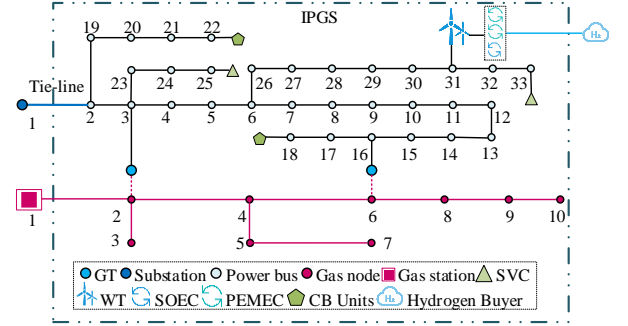


Fig. 4 Test System

To evaluate the effectiveness of the proposed strategy, five cases are designed for comparison, as shown in Table 1. The dispatch results in five cases are presented in Fig. 5- Fig.7. The results of the maximum relaxation gap of the pipeline are $8.2\text{e}-5$, and its value is small enough, so the relaxed MISOCP model is regarded as exact. The details of the five cases are described as follows:

Case 1: The single-stage day-ahead deterministic dispatch strategy by using the day-ahead prediction data is formulated for IPGS. Besides, stochastic prediction errors are not considered.

Case 2: The two-stage deterministic dispatch strategy with the latest intraday rolling prediction data by using deterministic MPC is formulated. Besides, stochastic prediction errors are not considered.

Case 3: The two-stage dispatch strategy with the latest intraday rolling prediction data by using stochastic MPC is formulated. Besides, stochastic prediction error are considered to follow traditional Gaussian distribution.

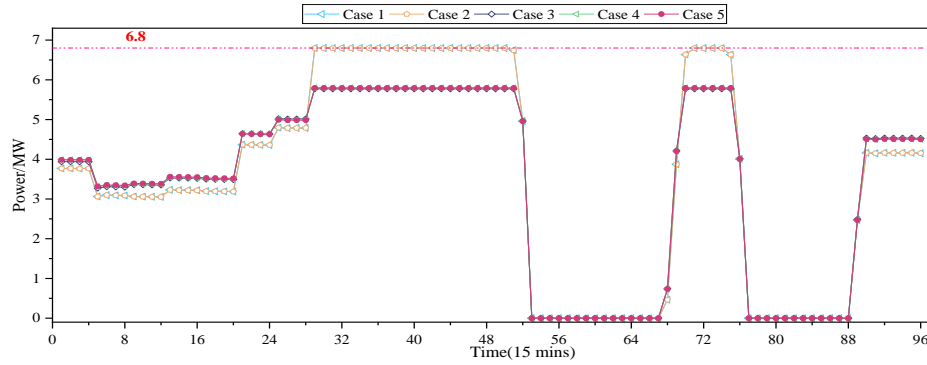
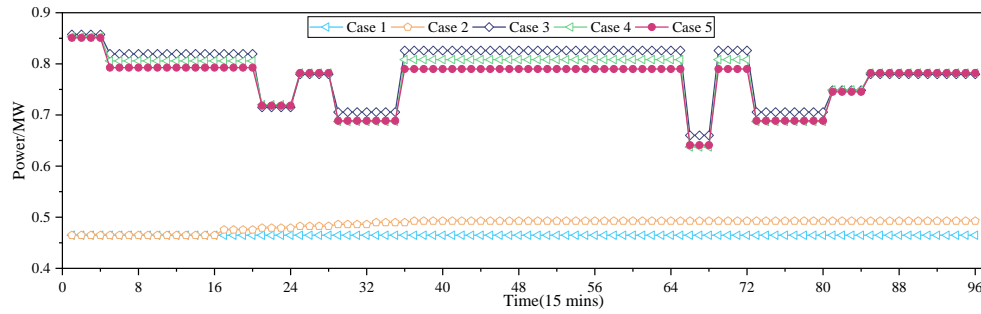
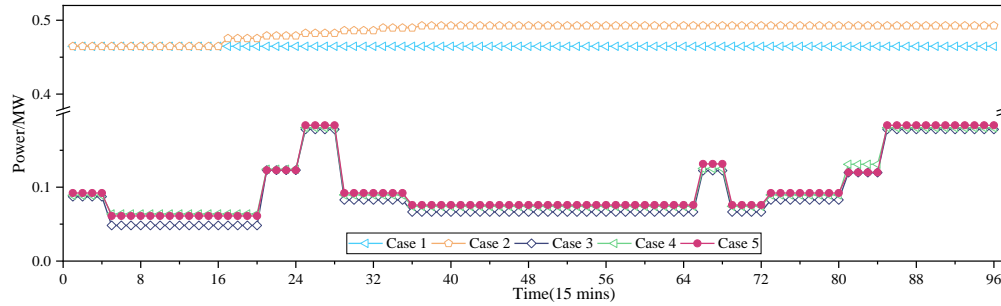
Case 4: The two-stage dispatch strategy with the latest intraday rolling prediction data by using deterministic MPC is formulated. Besides, the traditional GMM method is utilized to fit the distribution of stochastic prediction error.

Case 5: The two-stage dispatch strategy with the latest intraday rolling prediction data by using stochastic MPC is formulated. Besides, the VBGMM method is utilized to fit the distribution of stochastic prediction errors.

Table 1

Five Different Cases

						2	×	√	×	√	×	×
						3	×	√	√	√	×	×
						4	×	√	√	×	√	×
						5	×	√	√	×	×	√
Case	Single-Stage	Two-Stage	Prediction Errors	Gaussian	GMM	VBGMM						
1	√	×	×	√	×	×						

**Fig. 5** Dispatch results of the tie-line power in five cases**Fig. 6** Dispatch results of total upward reserve capacity of the gas-fired generators in five cases**Fig. 7** Dispatch results of total downward reserve capacity of the gas-fired generators in five cases**Table 2**

Comparison of Dispatch Results in Five Cases

Case	Wind Power (MWh)		Electric Load (MWh)		Operation Costs (\$)			Overload Times
	Predicted	Actual	Predicted	Actual	Budgeted	Adjustment	Actual	
1	32.28	24.06	166.81	177.39	34397.742	3569.91	37967.652	29
2	27.65	24.06	172.05	177.39	35821.612	1691.84	37513.452	29
3	27.65	24.06	172.05	177.39	35738.006	1361.19	37099.196	2
4	27.65	24.06	172.05	177.39	35733.478	1367.01	37100.488	2
5	27.65	24.06	172.05	177.39	35722.529	1366.00	37088.529	2

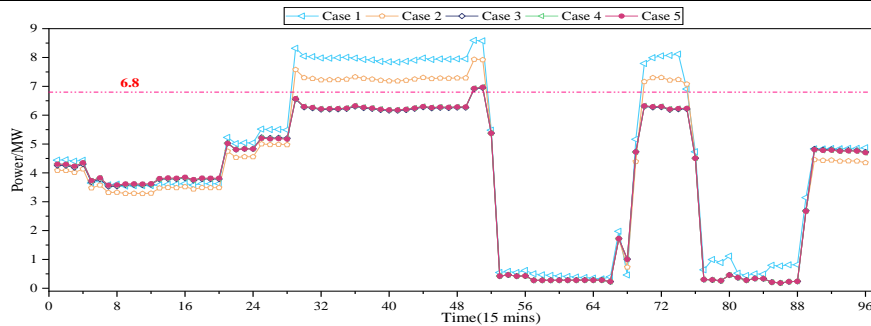


Fig. 8 Active power of the tie-line in five cases

5.2. Comparison of costs and times of overload

The comparison of dispatch results in five cases on a typical day is shown in Table 2. The day-ahead predicted wind power in case 1 is larger than the intraday value and actual value. Moreover, the day-ahead predicted value of electric load is lower than its intraday value and actual value. As a result, the actual value of prediction errors in day-ahead dispatch stage is the largest. Compared with single-stage dispatch (case 1), the latest intraday rolling predicted wind power and electric load in case 2 to case 5 are closer to the actual value, so the value of prediction errors is lower. Wind power costs are assumed to be zero in many references [8,11]. The budgeted operation costs in single-stage day-ahead dispatch (case 1) are lower because the predicted wind power is overestimated and assumed to be fully utilized to supply electric load, and the electric load is underestimated. Similarly, the budgeted operation costs in two-stage dispatch by using deterministic MPC are larger because of the more accurate and lower value of intraday predicted wind power and electric load.

As shown in Table 2, the first two columns present that when the single-stage day-ahead dispatch is conducted, 32.28 MWh of wind power is predicted to be available, but only 24.06 MWh of wind power is actually available, meaning that there is an 8.22 MWh deficit of wind power. Besides, only 166.81 MWh of electric load is predicted to be met, but 177.39 MWh of electric load actually needs to be met, meaning that there is 10.58 MWh surplus electric load. The total prediction errors in single-stage day-ahead dispatch are 18.8 MWh.

As the single-stage day-ahead deterministic dispatch does not consider the potential adverse effect of the stochastic prediction errors, the tie-line power in the 9th to 13th hour (15-min interval 29-52), and 18th hour to 19th hour (15-min interval 73-76) is very close to the maximum permissible value of the active power of the tie-line (6.8 MW). To cover the deficit of wind power and the surplus of electric load, multi-energy operator needs to purchase more power from the external power grid through the tie-line, resulting in the actual power of the tie-line exceeding its maximum permissible value (6.8 MW), as shown in Fig.8. The tie-line experienced overload 29 times in case 1.

As for the two-stage dispatch, 27.65 MWh of wind power is predicted to be available, but only 24.06 MWh of wind power actually is available, meaning that there is 3.59 MWh deficit of wind power. Similarly, 172.05 MWh of electric load is predicted to be met, but 177.39 MWh of electric load actually needs to be met, meaning that there is 5.34 MWh surplus of electric load. The value of prediction errors in two-stage dispatch is 8.93 MWh. The total amount of prediction errors is lower than the day-ahead one because of the more accurate intraday rolling prediction data of wind power and electric load. The two-stage dispatch by using the deterministic MPC in case 2 is similar to single-stage day-ahead one, but its actual overload power of the tie-line is lower, as shown in Fig.8. The total times of overload of the tie-line in case 2 is twenty-nine. By contrast, the dispatch by using the stochastic MPC in case 3 to case 5 considers the potential adverse effect of the stochastic

prediction errors, so its tie-line power in 8th to 13th, and 18th to 19th hour is lower. The actual power of tie-line is not easy to exceed the maximum permissible value (6.8 MW). The tie-line experienced overload twice in case 3 to case 5.

As mentioned above, the multi-energy operator needs to purchase more power from the external power grid through the tie-line to cover the deficit of wind power and meet the surplus of power demand and pay adjustment costs. In this work, the following assumptions about the adjustment price are made, that the adjustment price is the time of use price if the actual power of the tie-line does not exceed the maximum permissible value, otherwise, the adjustment price is the peak load price (230 \$/MWh). Then the adjustment costs are equal to the value of overload multiplies by the corresponding adjustment price. As a result, the actual operation costs are equal to the budgeted operation costs plus its adjustment costs. The total overload power in single-stage deterministic dispatch (case 1) is the largest, so its adjustment costs are 3569.91 \$, the largest in the five cases, then the actual costs, in this case, are the largest. Compared to the single-stage day-ahead deterministic dispatch, the total overload power in two-stage dispatch with intraday rolling prediction data by using deterministic MPC decreases obviously, so its adjustment costs and actual operation costs are also lower, illustrating the benefits of the two-stage dispatch by using MPC.

Then the adjustment costs and actual operation costs of the two-stage intraday dispatch by using stochastic MPC in case 3 to case 5 are lower than the adjustment costs and actual operation costs in case 2, illustrating the benefits of considering the uncertainty of stochastic prediction errors. Especially, compared to the Gaussian distribution in case 3 and GMM in case 4, the budgeted costs and actual operation costs with VBGM in case 5 are the lowest, illustrating the benefits of the data-driven chance constraints.

5.3. Comparison of reserve capacity

Many references only consider the uncertainty of electric load and set the reserve capacity according to the proportion of the maximum prediction value of electric load. As a result, the deterministic dispatch in case 1 and case 2 without considering the potential adverse effect of the prediction errors of wind power, and the multi-energy operator only sets the basic fixed load reserve (5 % of the maximum value of load), as shown in Fig.6 and Fig. 7. The total reserve capacity of the gas-fired generators in the first two cases cannot reflect the actual operation conditions and meet actual reserve capacity demand.

However, according to the data analysis of Elia in Section 5.2, the actual output of wind power tends to be lower and the actual demand for electric load tends to be larger. As a result, it is more reasonable for the multi-energy operator to set more upward reserve capacity and less downward reserve capacity. The single basic fixed load reserve capacity in case 1 and case 2 has two significant limitations, too radical for upward reserve capacity and too conservative for downward reserve capacity.

By contrast, the dispatch by using the stochastic MPC in case 3 to case 5 considers the potential adverse effect of the

stochastic prediction errors. The multi-energy operator set more upward reserve capacity and less downward capacity to mitigate the impact of these uncertainties. The reserve capacity in case 5 is more reasonable in the last three cases because it uses a data-driven CCSP method, and the benefits of this approach will be discussed in the next section 5.4.

5.4. Comparison of violation rate

The violation rate is a vital index to reflect and evaluate the quality of dispatch strategy in chance-constrained programming problems. The average violation rate is usually used to reflect and evaluate the quality of dispatch strategy, but it has a big defect, in that the extremely large or small violation rate is easily hidden. However, the maximum violation rate should be paid more attention to and alleviated by the multi-energy operator [36]. The IGPS with a high violation rate may not operate safely and economically. The average violation rate can result in a misleading interpretation of the performance of dispatch results and risk level. For example, the average violation rate of “1 % plus 9 %” and “4.9 % plus 5.1 %” are 5 %, but the former is worse if the tolerance level for risks is set as 5 %. More importantly, the large violation rate should be paid more attention to and eliminated. As shown in Table 3, the average violation of each chance constraint in the first two cases (deterministic dispatch) is the largest, about 20 %. Its value is too large, so the dispatch in the first two cases may not operate safely. The average violation in the last three cases is relatively small, showing the benefits of considering the uncertainty stochastic prediction errors. Specifically, the value of the average violation in the last three cases is very close, and the maximum value is only 1.08 times of the minimum value, so the numerical differentiation of the average violation is not obvious. More seriously, a little higher average violation does not mean it is not good. For instance, the average violation rate of “1 % plus 9 %” is 5 %, and the average violation rate of “4.9 % plus 5.15 %” is 5.025%. However, the latter one is definitely better in CCSP. The violation rate of overload in case 3 is too small, and its violation rate of insufficient downward reserve is too large, so its average violation rate is smaller. However, it should not think the dispatch results, in this case, are better. It is all the fault of the misleading interpretation of the average violation rate. Besides, the comparison results of operation costs in Table 2 also prove this point.

To evaluate the quality of the dispatch strategy and its robustness for the stochastic prediction errors better, a comprehensive evaluation index of the violation rate is proposed, which consists of two parts, the maximum violation rate and the cumulative value of the violation rate.

According to the definition of chance-constrained programming, the tolerance level α is a predefined value, which denotes the tolerance level of the multi-energy operator for the stochastic risks. Ideally, the violation rate should be less than or equal to the tolerance level. However, due to real stochastic data or inaccurate fitted probability density function, the actual violation rate may surpass the tolerance level. Still, an excessive violation rate exceeding the tolerance level (>5% in this paper) is undesirable. Therefore, the maximum violation rate is

proposed, which denotes the maximum value of the chance constraint in the whole dispatch period, as shown as follows:

$$MVR = \max(\alpha), \forall t = 1, \dots, T \quad (69)$$

where MVR denotes the maximum violation rate. α denotes the actual violation rate.

Similarly, the total violation rate of the above three chance constraints in the whole dispatch period should not be too large in the whole dispatch period (15 mins*96), so the cumulative value of violation rate (>5%) is proposed, as shown follows. The violation rate results based on the real historical data of the prediction errors are presented in Table 3.

$$CVR = \sum_{t=1}^T [(\bar{\alpha}_{sub,t}^{\max} - 5\%) + (\bar{\alpha}_{UR,t} - 5\%) + (\bar{\alpha}_{DR,t} - 5\%)] \quad (70)$$

where CVR denotes the cumulative value of the violation rate. $\bar{\alpha}_{sub,t}^{\max}$ denotes the maximum actual violation rate of overload of the tie-line power that exceeds 5% at time t . $\bar{\alpha}_{UR,t}$ and $\bar{\alpha}_{DR,t}$ denote the actual violation rate of insufficient upward/downward reserve that exceeds 5% at time t .

The cumulative value of the violation rate and maximum violation rate of insufficient upward reserve capacity in first two cases (deterministic dispatch) is the largest in five cases, as shown in Table. 3. The cumulative value of the violation of the two-stage dispatch by using deterministic MPC in case 2 is lower than the single-stage day-ahead deterministic dispatch, illustrating its benefits. The results of the cumulative value of the violation rate also show, that the dispatch results in case 1 and case 2 may not operate normally and safely in reality as their maximum violation rate and the cumulative value of the violation rate are too large and their robustness for the stochastic prediction errors is weak.

Compared to the above deterministic MPC in case 2, the cumulative value of violation rate and maximum violation rate considering the uncertainty in case 3 to case 5 decreases obviously, are lower than those value in case 2, illustrating the benefits of stochastic MPC.

The actual maximum violation rate of overload of the tie-line power of the Gaussian distribution in case 3 is 4.61%, much less than the predefined tolerance level (5%), which indicates the fitted probability density function with this method is inaccurate and the chance constraints of the tie-line, in this case, are more conservative than the actual requirement (5%). Besides, the actual maximum violation rate of insufficient upward reserve of the Gaussian distribution in case 3 is 5.99 %, more than the predefined tolerance level (5%), which indicates the fitted probability density function with this method is inaccurate and the chance constraints of upward reserve capacity in this case are more radical than the actual requirement. Similar conclusions can be drawn in case 4 (GMM method), and the cumulative value of the violation rate decreases because of the more accurate fitted probability density function.

In contrast, the maximum violation rate of overload and insufficient upward/downward reserve is relatively low and reasonable, very close to the predefined tolerance level (5%). Besides, the cumulative value of the violation rate is the lowest. The low maximum violation rate and the cumulative value of

the violation rate also indicate the fitting effect of the VBGMM method is accurate, and it exhibits the strongest robustness among all cases against stochastic prediction errors.

Compared to the above average violation rate, the numerical differentiation of the cumulative value of the violation rate in the last three cases is more obvious. More importantly, the more

concerned index by the multi-energy operator, the total violation level that exceeds the tolerance level during the entire dispatch period (24 hours), is more clearly presented by the proposed index in this paper.

Table 3
Comparison of Violation Rate in Five Cases

Case	Overload	Maximum Violation Rate		Average Violation Rate	Cumulative Value of Violation Rate (>5%) (Lower is Better)
		Insufficient Upward Reserve	Insufficient Downward reserve		
1	89.53%	39.67%	0.88%	20.31%	56.776
2	89.53%	39.67%	0.64%	19.43%	49.549
3	4.61%	5.14%	5.99%	3.63%	0.587
4	5.36%	5.36%	5.99%	3.81%	0.509
5	4.99%	5.24%	5.86%	3.94%	0.344

6. Conclusions

This work presents a two-stage data-driven dispatch strategy by using stochastic MPC. The key findings can be summarized as follows:

(1) Compared to the single-stage day-ahead dispatch, the actual operation costs and cumulative value of violation rate of the two-stage dispatch by using deterministic MPC are lower. This is due to the more accurate latest intraday rolling prediction data utilized in the dispatch strategy.

(2) The use of stochastic MPC in the two-stage dispatch leads to a significant decrease in actual operation costs and violation rates. This is because stochastic MPC takes into account the potential adverse effects of stochastic prediction errors.

(3) The proposed comprehensive evaluation index of the violation rate offers a clearer evaluation and presentation of the violation rate in five cases compared to the traditional single average violation rate index.

(4) The data-driven dispatch strategy by using stochastic MPC can reduce the potential adverse effect of stochastic prediction error. Furthermore, the fitting effect of the VBGMM method is the best, resulting in the lowest actual operation costs and cumulative violation rate values.

Acknowledgment

This work was supported by National Natural Science Foundation of China (51837004, U2066601).

References

- [1] Davidson DJ. Exnovating for a renewable energy transition. *Nat Energy* 2019;4:254–6. <https://doi.org/10.1038/s41560-019-0369-3>.
- [2] Xu S. The paradox of the energy revolution in China: A socio-technical transition perspective. *Renewable and Sustainable Energy Reviews* 2021;137:110469. <https://doi.org/10.1016/j.rser.2020.110469>.
- [3] Chen JJ, Zhuang YB, Li YZ, Wang P, Zhao YL, Zhang CS. Risk-aware short term hydro-wind-thermal scheduling using a probability interval optimization model. *Applied Energy* 2017;189:534–54. <https://doi.org/10.1016/j.apenergy.2016.12.031>.
- [4] Jiao PH, Chen JJ, Peng K, Zhao YL, Xin KF. Multi-objective mean-semi-entropy model for optimal standalone micro-grid planning with uncertain renewable energy resources. *Energy* 2020;191:116497. <https://doi.org/10.1016/j.energy.2019.116497>.
- [5] Hu X, Zhang H, Chen D, Li Y, Wang L, Zhang F, et al. Multi-objective planning for integrated energy systems considering both exergy efficiency and economy. *Energy* 2020;197:117155. <https://doi.org/10.1016/j.energy.2020.117155>.
- [6] Li Y, Li Z, Wen F, Shahidehpour M. Privacy-Preserving Optimal Dispatch for an Integrated Power Distribution and Natural Gas System in Networked Energy Hubs. *IEEE Trans Sustain Energy* 2019;10:2028–38. <https://doi.org/10.1109/TSTE.2018.2877586>.
- [7] Raheli E, Wu Q, Zhang M, Wen C. Optimal coordinated operation of integrated natural gas and electric power systems: A review of modeling and solution methods. *Renewable and Sustainable Energy Reviews* 2021;145:111134. <https://doi.org/10.1016/j.rser.2021.111134>.
- [8] Qi F, Shahidehpour M, Wen F, Li Z, He Y, Yan M. Decentralized Privacy-Preserving Operation of Multi-Area Integrated Electricity and Natural Gas Systems With Renewable Energy Resources. *IEEE Trans Sustain Energy* 2020;11:1785–96. <https://doi.org/10.1109/TSTE.2019.2940624>.
- [9] A. Mansouri M, Sioshansi R. Using Interim Recommitment to Reduce the Operational-cost Impacts of Wind Uncertainty. *Journal of Modern Power Systems and Clean Energy* 2022;10:839–49. <https://doi.org/10.35833/MPCE.2021.000573>.
- [10] Yang Y, Wu W, Wang B, Li M. Analytical Reformulation for Stochastic Unit Commitment Considering Wind Power Uncertainty With Gaussian Mixture Model. *IEEE Trans Power Syst* 2020;35:2769–82. <https://doi.org/10.1109/TPWRS.2019.2960389>.
- [11] Wang Z, Shen C, Liu F, Wu X, Liu C-C, Gao F. Chance-Constrained Economic Dispatch With Non-Gaussian Correlated Wind Power Uncertainty. *IEEE Trans Power Syst* 2017;32:4880–93. <https://doi.org/10.1109/TPWRS.2017.2672750>.
- [12] Wang Y, Dong W, Yang Q. Multi-stage optimal energy management of multi-energy microgrid in deregulated electricity markets. *Applied Energy* 2022;310:118528. <https://doi.org/10.1016/j.apenergy.2022.118528>.
- [13] Fang X, Wang Y, Dong W, Yang Q, Sun S. Optimal energy management of multiple electricity-hydrogen integrated charging stations. *Energy* 2023;262:125624. <https://doi.org/10.1016/j.energy.2022.125624>.
- [14] Mesbah A. Stochastic Model Predictive Control: An Overview and Perspectives for Future Research. *IEEE Control Syst* 2016;36:30–44. <https://doi.org/10.1109/MCS.2016.2602087>.
- [15] Li Z, Xu Y. Optimal coordinated energy dispatch of a multi-energy microgrid in grid-connected and islanded modes. *Applied Energy* 2018;210:974–86. <https://doi.org/10.1016/j.apenergy.2017.08.197>.
- [16] Jiang Y, Wan C, Wang J, Song Y, Dong ZY. Stochastic Receding Horizon Control of Active Distribution Networks With Distributed Renewables. *IEEE Trans Power Syst* 2019;34:1325–41. <https://doi.org/10.1109/TPWRS.2018.2879451>.
- [17] Ye L, Zhang Y, Zhang C, Lu P, Zhao Y, He B. Combined Gaussian Mixture Model and cumulants for probabilistic power flow calculation of integrated wind power network. *Computers & Electrical Engineering* 2019;74:117–29. <https://doi.org/10.1016/j.compeleceng.2019.01.010>.
- [18] Wan C, Cui W, Song Y. Probabilistic Forecasting for Power Systems With Renewable Energy Sources: Basic Concepts and Mathematical Principles. *Proceedings of the CSEE*. 2021; 19: 6493-6508. 10.13334/j.0258-

- 8013.pcsee.210931.
- [19] Sun W, Zamani M, Hesamzadeh MR, Zhang H-T. Data-Driven Probabilistic Optimal Power Flow With Nonparametric Bayesian Modeling and Inference. *IEEE Trans Smart Grid* 2020;11:1077–90. <https://doi.org/10.1109/TSG.2019.2931160>.
 - [20] Scikit-learn. (2022, Feb.). Variational Bayesian Gaussian Mixture. [Online]. Available: <https://scikit-learn.org/stable/modules/mixture.html#variational-bayesian-gaussian-mixture>
 - [21] National Grid. <https://www.nationalgrid.com/about-us>.
 - [22] Sun X, Qiu J. Hierarchical Voltage Control Strategy in Distribution Networks Considering Customized Charging Navigation of Electric Vehicles. *IEEE Trans Smart Grid* 2021;12:4752–64. <https://doi.org/10.1109/TSG.2021.3094891>.
 - [23] Ren J, Gu W, Wang Y, Ji W, Liu H, Cao G. Multi-time Scale Active and Reactive Power Coordinated Optimal Dispatch in Active Distribution Network Based on Model Predictive Control. *Proceeding of CSEE* 2018; 5. 1397-1407.
 - [24] Jiang Y, Xu J, Sun Y, Wei C, Wang J, Liao S, et al. Coordinated operation of gas-electricity integrated distribution system with multi-CCHP and distributed renewable energy sources. *Applied Energy* 2018;211:237–48. <https://doi.org/10.1016/j.apenergy.2017.10.128>.
 - [25] Blei DM, Jordan MI. Variational inference for Dirichlet process mixtures. *Bayesian Anal* 2006;1. <https://doi.org/10.1214/06-BA104>.
 - [26] Li Y, Schofield E, Gönen M. A tutorial on Dirichlet process mixture modeling. *Journal of Mathematical Psychology* 2019;91:128–44. <https://doi.org/10.1016/j.jmp.2019.04.004>.
 - [27] Li P, Ji H, Wang C, Zhao J, Song G, Ding F, et al. Coordinated Control Method of Voltage and Reactive Power for Active Distribution Networks Based on Soft Open Point. *IEEE Trans Sustain Energy* 2017;8:1430–42. <https://doi.org/10.1109/TSTE.2017.2686009>.
 - [28] Correa-Posada CM, Sanchez-Martin P. Integrated Power and Natural Gas Model for Energy Adequacy in Short-Term Operation. *IEEE Trans Power Syst* 2015;30:3347–55. <https://doi.org/10.1109/TPWRS.2014.2372013>.
 - [29] MATLAB. Fzero function 2022. [Online]. Available: <https://www.mathworks.cn/help/matlab/ref/fzero.html>.
 - [30] Wen Y, Qu X, Li W, Liu X, Ye X. Synergistic Operation of Electricity and Natural Gas Networks via ADMM. *IEEE Trans Smart Grid* 2018;9:4555–65. <https://doi.org/10.1109/TSG.2017.2663380>.
 - [31] Li Y, Li Z, Wen F, Shahidehpour M. Minimax-Regret Robust Co-Optimization for Enhancing the Resilience of Integrated Power Distribution and Natural Gas Systems. *IEEE Trans Sustain Energy* 2020;11:61–71. <https://doi.org/10.1109/TSTE.2018.2883718>.
 - [32] Dunning I, Huchette J, Lubin M. JuMP: A Modeling Language for Mathematical Optimization. *SIAM Rev* 2017;59:295–320. <https://doi.org/10.1137/15M1020575>.
 - [33] Baran M E, Wu F F. Network reconfiguration in distribution systems for loss reduction and load balancing[J]. *IEEE Power Engineering Review*, 1989, 9(4): 101-102.
 - [34] Jiang Y. Natural Gas and Power Integrated Regional Energy System Coordinated Operation Strategy. Wuhan University. 2019.
 - [35] Elia. (2021, Feb.). Wind power generation. [Online]. Available: <https://www.elia.be/en/grid-data/power-generation/wind-power-generation>.
 - [36] Cheng T. Exploiting Flexibility of Integrated Demand Response to Alleviate Power Flow Violation During Line Tripping Contingency. *JOURNAL OF MODERN POWER SYSTEMS AND CLEAN ENERGY* n.d.