Applied Energy 376 (2024) 124222

Contents lists available at ScienceDirect

Applied Energy

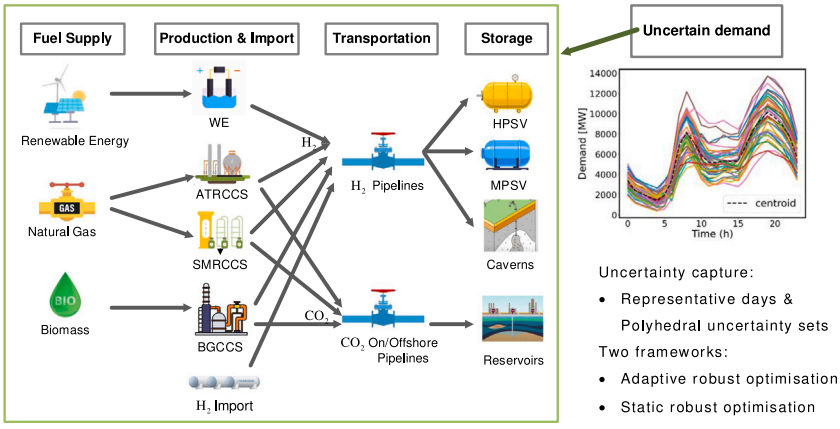
journal homepage: www.elsevier.com/locate/apenergy

Data-driven robust optimisation of hydrogen infrastructure planning under demand uncertainty using a hybrid decomposition method

Xu Zhou, Margarita E. Efthymiadou, Lazaros G. Papageorgiou, Vassilis M. Charitopoulos *

Department of Chemical Engineering, Sargent Centre for Process Systems Engineering, University College London (UCL), Torrington Place, WC1E 7JE, London, United Kingdom

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Hydrogen infrastructure planning
Net-zero
Heat decarbonisation
Data-driven robust optimisation
Column-and-constraint generation

ABSTRACT

In the race towards "Net-zero", hydrogen has emerged as one of the key alternatives to carbon-based fossil fuels for a sustainable decarbonisation. This work studies the spatially explicit multi-period hydrogen infrastructure planning under demand uncertainty that contributes to the heat decarbonisation in Great Britain. Demand uncertainty surrounding future hydrogen supply chains poses challenges to cost optimisation and system security, so uncertainty-resilient policies are required to ensure robust operations. In this work, we employ data-driven robust optimisation to develop a framework for uncertainty-aware representative days explicitly characterised by polyhedral uncertainty sets. The proposed framework is applied on a multi-period mixed-integer linear model with dual temporal resolution which aims to determine the optimal yearly investment decisions and hourly operational decisions for the hydrogen infrastructure planning under demand uncertainty. To efficiently solve the large-scale two-stage adaptive robust optimisation problem, a hybrid decomposition algorithm is developed based on a two-step hierarchical procedure and the column-and-constraint generation method, which can significantly reduce the computational complexity. The optimisation results highlight how uncertainty can result in the total cost increase, and verify the advantages on controlling solution conservatism in the adaptive robust optimisation compared to the static robust optimisation.

* Corresponding author.
E-mail address: v.charitopoulos@ucl.ac.uk (V.M. Charitopoulos).

https://doi.org/10.1016/j.apenergy.2024.124222
Received 23 May 2024; Received in revised form 9 July 2024; Accepted 13 August 2024
Available online 31 August 2024

Nomenclature

Abbreviations

SMRCCS	Steam Methane Reforming with Carbon Capture and Storage
ATRCCS	Autothermal Reforming with Carbon Capture and Storage
BGCCS	Biomass Gasification with Carbon Capture and Storage
WE	Water Electrolysis
CCG	Column-and-Constraint Generation
BCD	Block Coordinate Descent

Indices and sets

$g, \mathcal{G}$	index and set for region
$t, \mathcal{T}$	index and set for year period
$c, \mathcal{C}$	index and set for day (cluster)
$h, \mathcal{H}$	index and set for hour
$p, \mathcal{P}$	index and set for production technology
$r, \mathcal{R}$	index and set for reservoir
$s, \mathcal{S}$	index and set for storage technology
sa	index for storage cavern
$e, \mathcal{E}$	index and set for renewable technology
$Gimp_g$	set of regions in which international import are feasible
$\mathcal{N}_{gg'}$	set of total connections between neighbouring regions
$\mathcal{N}^{\text{pipe}}_{gg'}$	set of neighbouring regions that are connectable via H_2 pipelines
GS_{gs}	sets of region g in which storage technology s is located
GR_{gr}	connection sets of collection points g and reservoir r

Parameters

δ	ratio of hydrogen regional pipeline operating costs to capital costs (%)
$\overline{\delta}, \overline{\overline{\delta}}$	ratio of onshore/offshore CO_2 regional pipeline operating costs to capital costs (%)
η_{pt}	efficiency of production technology p at each time period t (MW H_2/MWh)
η^{we}_t	efficiency of WE technology at t (%)
τ	duration of time periods (years)
AV_{egch}	availability of renewable technology e in region g, cluster c and hour h (%)
ba_{gt}	maximum threshold of biomass consumption in region g and year t (MWh)
cap^{P}_p	unit capacity for production type p (MW/unit)
$cap^{\max}_p$	maximum capacity of a hydrogen production plant of type p (MW/unit)
cap^{R}_r	unit capacity of reservoir r (kg CO_2)
cap^{S}_s	unit capacity for storage type s (MWh/unit)
$cap^{\min}_s / cap^{\max}_s$	minimum/maximum capacity of storage technology s (MWh/unit)
c^p_t	fuel (biomass/gas) price at year t (£/MWh)
cl	power curtailment coefficient of renewables (%)
crf	capital recovery factor
ct_t	carbon tax at t (£/kg CO_2)
dfc_t / dfo_t	discount factor for capital/operating costs at t
$D^{\text{Pipe}}_{gg'}$	pipeline transmission distance from region g to region g' (km)
$D^{st}_{g,sa}$	pipeline transmission distance from region g to storage cavern sa (km)
D^{res}_{gr}	pipeline transmission distance from CO_2 collection point in region g to reservoir r (km)
et_t	CO_2 emissions target at t ($MtCO_2$)
I^{Up}_{gtc}	upper bound of hydrogen import rate at g, t, c (MW)
la_{eg}	available capacity upper bound of renewable technology e at region g (MW)
LT^{pipe}	lifetime of pipelines (year)
LTP_p	lifetime of production technology p (year)
LTS_s	lifetime of storage technology s (year)
p^{imp}	price of hydrogen import (£/MW)
pc	capital cost of hydrogen pipelines (£/km)
$\overline{pc}/\overline{\overline{pc}}$	capital cost of onshore/offshore CO_2 pipelines (£/km)
pcc_{pt}	capital cost of production technology p at year t (£/MW)
$poc^{\text{F}}_{pt} / poc^{\text{V}}_{pt}$	fixed/variable operating production cost of p at year t (£/MW)
q^{Hmax}	maximum flowrate of H_2 pipelines (MW)
q^{Cmax}	maximum flowrate of CO_2 pipelines (kg CO_2/h)
$Q^{\text{Imax}}_s / Q^{\text{Rmax}}_s$	maximum storage/withdrawal rate for each storage type s (MW)
rc_{et} / ro_{et}	capital/operational cost of renewable technology e at t (£/MW)
RD_p / RU_p	ramp-down/ramp-up rate of production technology p (%)
scc_{st}	capital cost of storage unit s at t (£/MWh H_2)
soc^{F}_{st}	fixed operating production cost of storage technology s at t (£/MWh H_2/y)
soc^{V}_{st}	variable operating production cost of storage technology s at t (£/MW)
WF_c	weight of cluster c (day)

1. Introduction

To be in line with the current low-carbon emission goals, hydrogen as an alternative energy carrier to natural gas has increasingly attracted global attention, demonstrating its potential in contributing to sustainable decarbonisation energy solutions [1–3]. In Great Britain (GB), the heat sector accounts for approximately one-third of the total national CO_2 emissions [4,5]. Hydrogen has been progressively recognised as a crucial vector towards the heat decarbonisation pathways in GB [6]. The reports from Committee on Climate Change, National Grid and Department for Business, Energy & Industrial Strategy (BEIS) suggest that hydrogen demand will increase significantly by early 2030 and the 7–20 GW of production capacity may be needed by 2035 [7,8]. Therefore, there is an urgent need to improve and expand hydrogen energy infrastructure networks to satisfy the increasing demand.

The large-scale hydrogen infrastructure planning problems on a nationwide scale have been explored by the research community within

y^{c}_{pt}/y^{e}_{pt}	coefficient of CO_2 capture/emissions for production technology p at t (kg CO_2/MWh H_2)

Integer Variables

IP_{pgt}	total investment number of new production technology p at g,t (units)
NP_{pgt}	total number of available production technology p at g,t (units)
IS_{sgt}	total investment number of new storage facility s at g,t (units)
NS_{sgt}	total number of available storage facility s at g,t (units)

Binary Variables

$AY_{gg't}$	availability of hydrogen pipeline between regions g and g' at t
$AY^{S}_{g,sa,t}$	availability of hydrogen pipeline from region g to storage cavern sa at t
$\overline{AY}_{gg't}$	availability of onshore CO_2 pipeline between regions g and g' at t
$\overline{\overline{AY}}_{grt}$	availability of offshore CO_2 pipeline from region g to reservoir r at t
$Y_{gg't}$	establishment of hydrogen pipeline between regions g and g' at t
$Y^{S}_{g,sa,t}$	establishment of hydrogen pipeline from region g to storage cavern sa at t
$\overline{Y}_{gg't}$	establishment of onshore CO_2 pipeline between regions g and g' at t
$\overline{\overline{Y}}_{grt}$	establishment of offshore CO_2 pipeline from region g to reservoir r at t

Continuous Variables

CL_{gtch}	renewable power curtailment at g,t,c,h (MW)
D_{gtch}	uncertain hydrogen demand at g,t,c,h (MW)
E_t	total CO_2 emissions at t ($MtCO_2$)
I_{gtch}	hydrogen flowrate of international import at g,t,c,h (MW)
IR_{egt}	new invested capacity of renewable technology e at g,t (MW)
NR_{egt}	available capacity of renewable technology e at g,t (MW)
Pr_{pgtch}	hydrogen production rate of production technology p at g,t,c,h (MW)
$\widehat{Pr}_{egtch}$	electricity production from renewable technology e at g,t,c,h (MW)
$Q_{gg'tch}$	hydrogen flowrate via pipeline from region g to g' at t,c,h (MW)
Q^{I}_{gstch}	hydrogen flowrate via pipeline from region g to storage s at t,c,h (MW)
Q^{R}_{sgtch}	hydrogen flowrate via pipeline from storage s to region g at t,c,h (MW)
$\overline{Q}_{gg'tch}$	CO_2 flowrate via onshore pipeline from region g to g' at t,c,h (kg CO_2/h)
$\overline{\overline{Q}}_{grtch}$	CO_2 flowrate via offshore pipeline from a collection point in region g to reservoir r at t,c,h (kg CO_2/h)
RI_{rt}	CO_2 storage state of reservoir r at t (kg CO_2)
St_{sgtch}	hydrogen storage state of storage facility s at g,t,c,h (MWh)
V^{bio}_{gt}	total biomass consumption at g,t (MWh)
TC	total cost (£)
PCC/POC	production capital/operational cost (£)
SCC/SOC	storage capital/operational cost (£)
$PLCC$	pipeline transportation capital cost (£)
$PLOC$	pipeline transportation operational cost (£)
CEC	carbon emissions cost (£)
IIC	hydrogen import cost (£)
ReC	total capital and operational cost of renewable technologies (£)
FC	total fuel cost for natural gas and biomass consumption (£)

the context of "Net-zero" agenda by 2050 [9]. In [10], a spatially-explicit multi-period MILP model was developed for the design of hydrogen supply chain (HSC) in the UK involving the range of production technologies, scales, transportation modes and carbon capture and storage elements. [11] aimed to soft-link resource-technology network and whole electricity system investment model with different geographical scopes and temporal resolutions involving the interaction between electricity and hydrogen infrastructure for decarbonising GB heat supply by 2050. The study in [12] focused on the design of nationwide H_2 and CO_2 infrastructure using a spatio-temporal based model to determine the optimal choices of hydrogen production, transportation and storage technologies towards the heat decarbonisation of GB. [13] proposed a model to design and analyse national HSC with spatial resolution in Germany based on different technologies options for storage and transport. Two long-term, multi-regional and monthly time-scale MILP optimisation models of China natural gas supply system and HSC were established in [14] from 2020 to 2060. [15] developed a generalised framework for co-optimising infrastructure investments across the electricity and HSC considering the spatio-temporal variations in energy demand and supply, and applied it to the U.S. Northeast region. [16] presented a multi-objective MILP optimisation model for HSC network to determine the locations and sizes of hydrogen facilities, production technology, transportation unit and distribution route. This model was applied to the state of Texas. A novel mixed-integer nonlinear programming modelling and optimisation framework related to the design and long-term capacity expansion of HSC was presented in [17] providing an optimal roadmap for the 10-year plan of hydrogen infrastructure development in the state of California. [18] studied the design of HSC in Turkey to meet the hydrogen demand from 2021 to 2050 by using the mixed integer programming model. The multi-objective optimisation models were proposed in [19,20] for the HSC design considering the production, storage, transportation, and distribution, and applied the models to the case in southern France and Hungary, respectively. [21] addressed the optimal design of sustainable HSC for vehicle use in the UK, which was formulated as an MILP problem to determine the optimal combination of different technologies, their capacities and location, and the transport flows. The work in [22] analysed how variations in renewable and Power-to-X technology costs can affect the optimal solution including the optimal technology capacity and the role of hydrogen under the hourly resolution in a long-term energy planning setting for the Italian energy system decarbonisation pathway. [23] presented a spatially and temporally-explicit multi-period supply chain model for three pathways (electricity, heat, hydrogen) in decarbonising the UK. Our previous work [24] presented an MILP optimisation framework focusing on the

hydrogen infrastructure decisions with 5-year steps from 2035 to 2050 and hourly resolution for the transition of heat sector in GB.

Although the aforementioned works and many existing works in the literature [25] have mathematically modelled various forms of HSC problems and presented effective outcomes, they assumed the deterministic models with fixed parameters. Such models, however, fail to consider the uncertainties inherent to underlying HSC problems, such as the uncertain hydrogen demand and renewable energy generation, which may render the solution to the deterministic model infeasible when model parameters fluctuate. This could lead to substantial economic losses and pose a threat to the overall security of HSC [26,27].

To this end, many studies have explored stochastic programming (SP) [28] representing uncertainty factors as probabilistic scenarios and robust optimisation (RO) techniques considering the worst-case scenario realisation to hedge against the uncertainty [29]. However, the probability distribution of uncertainties needed in SP may be difficult to obtain in practice, and stochastic programming is known to lead to computational complex problems due to the rapid increase in problem size based on the numbers of scenarios needed to accurately capture the uncertainty. On the other hand, RO provides a trade-off between feasibility and computational tractability by establishing a deterministic uncertainty set that describes the variability of uncertain parameters. RO can be further categorised into static robust optimisation (SRO) where all decisions are made proactively, and adaptive robust optimisation (ARO) where part of decisions can be adjustable once the uncertainty is realised. Naturally, ARO has gradually become a more attractive alternative to SRO due to the less conservative solutions which however come with a considerable computational cost. RO techniques have been widely applied to energy systems problems such as the power network expansion planning problems [30–32], gas network such as the natural gas infrastructure planning [33], charging stations and integrated energy systems [34], among others [35].

Despite the plethora of research works focusing on uncertainties for the HSC planning, most of the studies addressed the demand uncertainties by using sample average approximation technique [36] and scenario-based approach in SP [37–43], while only a handful explored RO technique for risk-averse decision-making in HSC planning. [44] proposed a single-level robust hydrogenation network optimisation model under demand uncertainty considering the capacity expansion of hydrogen refuelling stations and flexible transportation modes between sites. In order to meet the hydrogen needs of chemical plants, a static robust model was used in [45] to determine the optimal capacity configuration of a PV/battery/hydrogen system considering the uncertainty of PV output and hydrogen demand. A two-stage production and replenishment problem was investigated in [46] for HSC considering the demand uncertainty and the risk of pipeline disruptions, which was formulated as a distributionally robust optimisation problem. A column and constraint generation (CCG) algorithm was further proposed to solve this RO problem. [47] investigated the distributionally RO problem under uncertainty to determine the optimal planning strategy for the cross-regional hydrogen energy storage systems. A two-stage robust design and planning problem for an integrated HSC under demand uncertainty described by a budget-based uncertainty set was addressed in [48] by using a benders decomposition based algorithm to eliminate the inherent spatiotemporal imbalance between renewable energy supply and demand. Nevertheless, this model did not consider the interregional pipeline transmission of hydrogen. Note that the spatial scale considered in [46–48] is just sub-regional rather than a nationwide scale. [49] developed an MILP model to configure the HSC network in an international scale considering strategic and tactical decisions on the type and location of hydrogen production/storage sites, distribution system and physical form of product delivery. The robust possibilistic programming method was used to deal with the hydrogen demand uncertainty. However, it only considered the single planning period, which ignored the multi-period nature of investment and operational decisions.

In this work, we focus on the large-scale hydrogen infrastructure planning problem for GB under demand uncertainty, which is formulated as a spatially-explicit, multi-period MILP model on a nationwide scale. Given their multi-scale nature, energy infrastructure planning models utilise time compression techniques to generate a set of representative days, aiming to achieve a balance between accuracy and complexity. [50] designed an LP model for the planning of highly renewable US power grid based on a representative subset of days, which deterministically solves for the least-cost portfolio of generators, storage, and transmission. For the case of electric power infrastructure planning, [51] selected representative days with hourly resolution from historical data via the clustering procedure to model each year, and proposed a decomposition algorithm based on nested benders decomposition to solve the deterministic multi-period MILP problem. The same authors further extended their model to consider uncertainty in [52] where operating uncertainties were captured through different scenarios of representative days' profiles, and solved the resulting large-scale problem using stochastic dual decomposition. However, the above studies and most of the existing literature mainly focused on power systems. The issue of systematically accounting for the uncertainty introduced through the deployment of representative days within energy infrastructure planning models remains largely unexplored. To this end, we introduce data-driven SRO and ARO models involving the construction of uncertainty sets to explicitly characterise potential demand uncertainty realisations. The polyhedral uncertainty set is adopted in this paper, which can capture correlations between high-dimensional uncertain parameters [53]. Note that this work does not consider the uncertainty of renewable energy generation (REG) because modelling it in a RO context can exacerbate unnecessarily the conservativeness of the approach due to the intrinsic volatility that needs to be captured in uncertainty sets. In addition, we have identified from our previous paper [54] that demand and REG are uncorrelated (the same as solar and wind) and hence it would be more appropriate for REG to be modelled independently using a scenario-based approach [55].

To the best of our knowledge, the spatially-explicit multi-period hydrogen infrastructure planning problems under uncertainty has not been studied through a RO approach. The contributions of this paper are summarised as follows:

(1) We investigate data-driven robust optimisation frameworks including both static and adaptive robust optimisation to explicitly account for the demand uncertainty for the large-scale hydrogen infrastructure planning over multi-spatial and temporal scales, resulting in efficient decision support under flexible and variable demand.
(2) The uncertainty is systematically captured through the deployment of uncertain representative days and the introduction of polyhedral uncertainty sets. This enables the consideration of a wide range of operational scenarios while maintaining computational feasibility, thereby reducing the likelihood of inappropriately selecting certain demand profiles.
(3) We propose a hybrid decomposition method under the CCG algorithm framework based on a two-step hierarchical procedure and block coordinate descent methods to solve the underlying two-stage adaptive robust optimisation problem, which can significantly reduce the computational time.
(4) The computational experiments on hydrogen infrastructure planning in GB over the 5-year steps 2035–2050 show the impact of uncertainty on the decision-making of investment and operation through a comparative analysis between the robust and the deterministic models. The computational results further verify the advantages on controlling conservatism of the adaptive robust optimisation compared to the static robust optimisation.

The rest of the paper is structured as follows. In Section 2, we give a formal problem statement, and in Section 3 we introduce the modelling strategies in spatial and temporal scales and how to capture the demand uncertainty. In Section 4, we formulate a data-driven static robust optimisation problem for hydrogen infrastructure planning and its solution strategy. In Section 5, a data-driven adaptive robust optimisation problem is presented, and a hybrid decomposition method is further proposed to handle this complex two-stage MILP problem. Results are presented and analysed in Section 6. Finally, we conclude this work in Section 7.

2. Problem statement

The proposed large-scale hydrogen infrastructure planning problem involves choosing the optimal investment strategy for each region in GB over a long-time horizon, and hourly operating decisions to meet the uncertain hydrogen demand.

The production, transmission, storage, and renewable technologies are given, and we consider (as shown in Fig. 1):

- Production technologies: Steam Methane Reforming with Carbon Capture and Storage (SMRCCS) & Autothermal Reforming with Carbon Capture and Storage (ATRCCS) that consume natural gas, Biomass Gasification with Carbon Capture and Storage (BGCCS), Water Electrolysis (WE).
- Transmission modes: Due to the substantial amount of hydrogen demand, the road transportation by trailer for hydrogen is impractical [12]. Consequently, we only focus on pipeline transport for H_2 and CO_2, which is the most efficient and widely used way.
- Storage technologies: H_2 storage involves Medium Pressure Storage Vessel (MPSV) and High Pressure Storage Vessel (HPSV) in each region, and cavern storage in a few specific regions. CO_2 can be stored by pipelines to reservoirs.
- Renewable technology: Solar, Wind Onshore and Wind Offshore.

Also given are:

- Capital and operational costs for different production technologies, storage sites, and pipelines of H_2 and CO_2;
- H_2 import price, carbon tax and capture rates for CO_2 emissions and the emission target;
- Regional connections, availability and distances for pipelines;
- Operating characteristics such as minimum/maximum capacity, ramp-up/ramp-down rates and lifetime of production plants and storage sites, maximum flowrate in pipelines, capacity of H_2 caverns & vessels and CO_2 reservoirs;
- Renewables availability in each region, year, day and hour;
- Historical data for gas-served heating demand in each region, year, day and hour.

To determine:

- Worst-case daily hydrogen demand over regions and years;
- Location and operating schedule for each production technology, storage site and renewable farm;
- Electricity generation of renewable energies;
- Investments of production and storage technology in each region, and transmission investments between regions over years for H_2 and CO_2;
- Hourly H_2 and CO_2 flowrates between regions over years;
- Hourly H_2 production, storage and import rates over regions and years to satisfy the hydrogen demand.

So as to: minimise the overall system costs subject to investment, operational and environmental constraints.

The main assumptions in this work are summarised below:

- Gas demand data are used as a proxy for gas consumption that is needed for domestic, commercial and industrial heating [6];

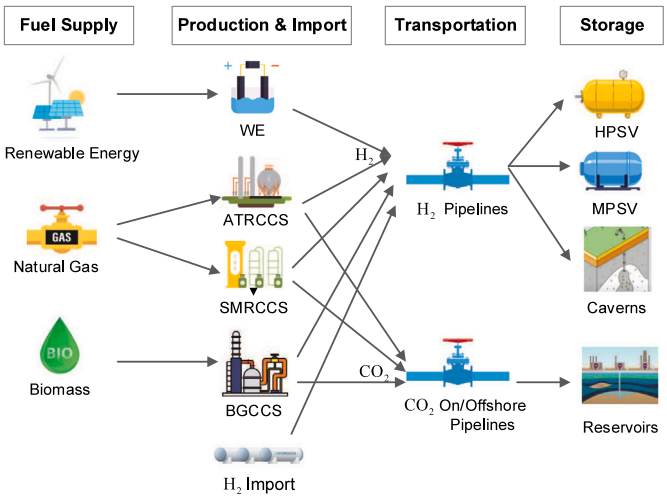


Fig. 1. The hydrogen supply chain.

- A gas price prediction is used for the 5-year step based on the National Grid's Future Energy Scenarios 2022 [7];
- All CO_2 and H_2 pipelines have identical diameters, respectively;
- The computation of transmission distances is based on the measurement between the centroids of each region;
- The cost and maximum capacity for each technology is deterministic;
- WE production technology is assumed to only consume electricity from the renewable energies;
- Hydrogen pipeline connections are configured based on the layout of the existing gas pipeline network in GB;
- The variable operational cost for renewable farms is assumed to be zero;
- The hydrogen distribution within each region and its export are not considered;
- We focus solely on gaseous hydrogen.

For the sake of brevity as the deterministic model that forms the basis is adopted by [24], the detailed equations are provided in the Appendix. Next we will delve into the reformulation of this model to incorporate uncertainty-aware representative days and its robust solution.

3. Modelling strategies under uncertainty

The multi-period and spatially explicit hydrogen infrastructure planning is formulated as an MILP optimisation problem in this work. To be able to solve it, it is crucial to reduce the problem scale so as to alleviate the computational complexity, which requires the modelling aggregations and approximations considering the multi-scale nature. In addition, the uncertainty of hydrogen demand can impact decision-making, potentially leading to significant economic losses and even compromising the security of energy systems. Therefore, how to capture the demand uncertainty when modelling is pivotal. In the next subsections, we will introduce the days clustering modelling method, the concept of representative days and polyhedral uncertainty sets.

3.1. Modelling representation in spatial and temporal scales

The hydrogen production, demand, and associated storage and transmission costs are influenced by geographical characteristics and existing infrastructure. Renewable energy availability and production also vary regionally. Therefore, it is essential to consider geographical factors in hydrogen infrastructure planning [56,57]. To this end, we

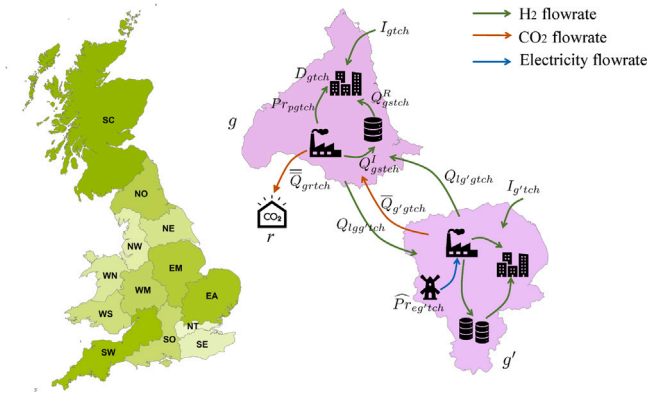


Fig. 2. Conceptual representation of regions and energy flows.

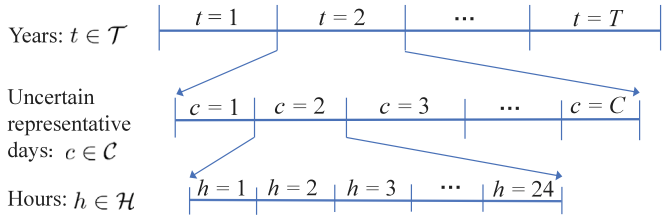


Fig. 3. Multi-scale representation of hydrogen infrastructure planning.

adopt a spatially explicit model wherein GB is disaggregated into 13 regions based on the established local distribution zones of gas network as shown in Fig. 2 (left) [6]. The 13 regions are denoted by the set $\mathcal{G}$. Fig. 2 (right) displays potential energy interactions for regions g and g'. Hydrogen can be generated by using renewable generation, and is transmitted from one region to other regions by pipelines transport if they are connected. The generated CO_2 will be stored as well by reservoirs to satisfy the emission target of GB.

With respect to the temporal scale, we consider not only a strategic 5-year investment horizon which is essential for infrastructure planning, but also the hourly operations because of hydrogen demand variability. To reduce the computational time, each year $t \in \mathcal{T}$ is modelled by using C representative days with hourly resolution (24 subperiods), which is widely adopted in the multi-scale MILP energy planning literature [54,58]. More specifically, each year's demand data are disaggregated into C clusters by using k-medoids clustering method, and each cluster $c \in \mathcal{C}$ encompasses a specific collection of hourly demand profiles [6]. As mentioned in Section 1, these representative days assume a deterministic hourly profile of their attributes which completely neglects the intra-cluster variability that can result in significant underestimation of the uncertainty the future infrastructure is bound to face. To overcome this issue, we apply a polyhedral uncertainty set to capture the demand uncertainty for each cluster. Thus, the representative hydrogen demand profiles are not deterministic and pre-specified anymore. Robust optimisation usually aims to minimise the cost under the worst case scenario, i.e., under worst daily demand profile of this uncertainty set. The number of days within each cluster c is viewed as the weight (WF_c) of representative day (also denoted by c). The multi-scale representation of the problem is more intuitively visualised in Fig. 3. Next, we will introduce how we establish the data-driven polyhedral uncertainty sets.

3.2. Data-drive robust representative days

Polyhedral uncertainty sets have been widely adopted to describe the range of uncertain parameters by a finite number of linear inequalities in robust optimisation [59]. They reduce computational complexity due to their linear nature and flexibility to capture uncertainty properties, like correlation.

Our point of departure is the method introduced in [53] to construct polyhedral uncertainty set from historical data by using principal component analysis (PCA) and Kernel density estimation (KDE). The uncertain demand data matrix for each region g, year t and cluster c is denoted by $\mathbf{X}_{gtc} = \left[\mathbf{D}_{gtc}^{(1)}, \ldots, \mathbf{D}_{gtc}^{(N_{gtc})}\right]^\top \in \mathbb{R}^{N_{gtc}\times 24}$ where each row represents an uncertainty demand data point with 24-dimensional (hourly) space, N_{gtc} is the total uncertainty data points, and $\mathbf{D}_{gtc} = [D_{gtc,1}, D_{gtc,2}, \ldots, D_{gtc,24}]^\top \in \mathbb{R}^{24}$ denotes daily hydrogen demand with hourly demand $D_{gtch} \in \mathbb{R}, h \in \{1, 2, \ldots, 24\}$.

Firstly, we standardise the data matrix $\mathbf{X}_{gtc}$ to zero-mean:

$$\mathbf{X}_{gtc}^0 = \mathbf{X}_{gtc} - \mathbf{e}\boldsymbol{\mu}_{gtc}^\top, \quad \forall g, t, c \tag{1}$$

where $\mathbf{X}_{gtc}^0$ is an uncertainty data matrix after standardisation, the averaged value $\boldsymbol{\mu}_{gtc} = \frac{1}{N_{gtc}}\sum_{i=1}^{N_{gtc}} \mathbf{D}_{gtc}^{(i)}$, and $\mathbf{e}$ is a unit column vector. Next, we perform a PCA to obtain orthogonal principal components via eigenvalue decomposition to the sample covariance matrix $\mathbf{S}_{gtc} = \frac{1}{N_{gtc}-1}\mathbf{X}_{gtc}^{0\top}\mathbf{X}_{gtc}^0$. Denote the eigenvectors of $\mathbf{S}_{gtc}$ by $\mathbf{P}_{gtc} = \left[P_{gtc,1}, \ldots, P_{gtc,24}\right] \in \mathbb{R}^{24\times 24}$, such that $\mathbf{S}_{gtc} = \mathbf{P}_{gtc}\Lambda_{gtc}\mathbf{P}_{gtc}^\top$ where Λ_{gtc} is the diagonal matrix containing all the eigenvalues. Each eigenvector corresponds to a principal component.

Then we project each uncertain demand data $\mathbf{D}_{gtc}^{(i)}$ onto each principal component $k = 1, \ldots, 24$:

$$\mathbf{t}_{gtc,k}^{(i)} = P_{gtc,k}^\top\left[\mathbf{D}_{gtc}^{(i)} - \boldsymbol{\mu}_{gtc}\right], \quad i = 1, \ldots, N_{gtc}, \quad \forall g, t, c \tag{2}$$

By using the KDE method on the projected data matrix $\mathbf{T}_{gtc,k} = \left[\mathbf{t}_{gtc,k}^{(1)}, \ldots, \mathbf{t}_{gtc,k}^{(N_{gtc})}\right]$, we can calculate the estimated probability density function $\hat{f}_{gtc,k}^{\mathrm{KDE}}(\xi_{gtc,k})$ for the latent uncertainty $\xi_{gtc,k}$ along the kth principal component, such that

$$\hat{f}_{gtc,k}^{\mathrm{KDE}}(\xi_{gtc,k}) = \frac{1}{N_{gtc}}\sum_{i=1}^{N_{gtc}} K_h\left(\xi_{gtc,k}, \mathbf{t}_{gtc,k}^{(i)}\right) \tag{3}$$

where K_h is a Gaussian kernel function with a bandwidth h. Denote the cumulative density function of $\xi_{gtc,k}$ by $\hat{F}_{gtc,k}^{\mathrm{KDE}}(\xi_{gtc,k})$. Its corresponding quantile function can be expressed as follows:

$$\widetilde{F}_{gtc,k}^{\mathrm{KDE}}(\alpha_{gtc}) = \min\left\{\xi_{gtc,k} \in \mathbb{R} \middle| \hat{F}_k^{\mathrm{KDE}}(\xi_{gtc,k}) \geq \alpha_{gtc}\right\} \tag{4}$$

where α_{gtc} is a predefined parameter related to confidence level.

Define $\underline{\xi}_{gtc} = \left[\widetilde{F}_{gtc,1}^{\mathrm{KDE}}(\alpha_{gtc}), \ldots, \widetilde{F}_{gtc,24}^{\mathrm{KDE}}(\alpha_{gtc})\right]^\top \in \mathbb{R}^{24}$ and $\overline{\xi}_{gtc} = \left[\widetilde{F}_{gtc,1}^{\mathrm{KDE}}(1-\alpha_{gtc}), \ldots, \widetilde{F}_{gtc,24}^{\mathrm{KDE}}(1-\alpha_{gtc})\right]^\top \in \mathbb{R}^{24}$. Finally, we get the following polyhedral uncertainty set D_{gtc}^{PUS} for each region g, year t and cluster c:

$$D_{gtc}^{\mathrm{PUS}} := \left\{\mathbf{D}_{gtc} \in \mathbb{R}^{24} \middle| \begin{array}{l} \mathbf{D}_{gtc} = \boldsymbol{\mu}_{gtc} + \mathbf{P}_{gtc}\cdot\xi_{gtc}, \\ \xi_{gtc} = \underline{\xi}_{gtc}\circ\mathbf{z}_{gtc}^- + \overline{\xi}_{gtc}\circ\mathbf{z}_{gtc}^+, \\ \mathbf{0} \leq \mathbf{z}_{gtc}^-, \mathbf{z}_{gtc}^+ \in \mathbb{R}^{24} \leq \mathbf{1}, \\ \mathbf{1}^\top\left(\mathbf{z}_{gtc}^- + \mathbf{z}_{gtc}^+\right) \leq \Phi_{gtc}, \\ \mathbf{z}_{gtc}^- + \mathbf{z}_{gtc}^+ \leq \mathbf{1} \end{array}\right\} \tag{5}$$

where $\boldsymbol{\mu}_{gtc}$ is the averaged value introduced in (1), $\mathbf{0}$ and $\mathbf{1}$ are column vectors of all zeros and ones, and $\mathbf{z}_{gtc}^-$ and $\mathbf{z}_{gtc}^+$ are the backward deviation and forward deviation vectors, respectively. The symbol $\circ$ denotes the Hadamard product. The lower and upper bound vectors $\underline{\xi}_{gtc}$ and $\overline{\xi}_{gtc}$ define the confidence interval of latent uncertainties. The parameter Φ_{gtc} is the uncertainty budget that controls the maximum extent to which latent uncertainties are allowed to deviate from their average values, i.e., controls the level of conservatism by changing

locations of hyperplane cuts (i.e., changing the size) of the uncertainty set. Note that Φ_{gtc} does not change the polyhedral structure of the uncertainty set.

In summary, in this section data-driven methods (Clustering, KDE and PCA) are used to construct polyhedral uncertainty sets from historical demand data when developing the representative days of the planning problem, which is fundamental to the next robust optimisation frameworks. The proposed uncertainty sets can flexibly capture compact regions of uncertainty in a nonparametric way and consider the data correlations between different hours by using PCA, thus enhancing the practicality and efficiency of the optimisation solution.

4. Data-driven static robust optimisation

In order to hedge against the hydrogen demand uncertainty, in this section, we develop a data-driven static robust framework and its solution for the hydrogen infrastructure planning involving the total system cost and corresponding constraints.

4.1. Static robust optimisation formulation

The total cost (TC) considered in the planning model consists of the production capital & operational cost (PCC) & (POC), storage capital & operational cost (SCC) & (SOC), pipeline transportation capital & operational cost ($PLCC$) & ($PLOC$), carbon emissions cost (CEC), hydrogen import costs (IIC), renewable cost (ReC), and total fuel cost (FC) for the natural gas and biomass consumption. To this end, we employ Eq. (6) as shown below.

$$\begin{aligned} TC = {} & PCC + SCC + POC + SOC + PLCC \\ & + PLOC + CEC + IIC + ReC + FC \end{aligned} \tag{6}$$

The detailed expression for each cost term can be found in Appendix (A.1) – (A.10).

The static robust MILP model for the hydrogen infrastructure planning problem is shown below, where the demand profiles of uncertain representative days belong to polyhedral uncertainty sets obtained in (5) of Section 3.2, allowing for a broader range of intra-day operational scenarios.

$$\begin{aligned} & \min\ TC \\ & \text{s.t. (6), (A.1) – (A.39),} \\ & \quad \mathbf{D}_{gtc} \in \mathcal{D}^{\mathrm{PUS}}_{gtc}, \quad \forall g \in \mathcal{G}, t \in \mathcal{T}, c \in C \end{aligned} \tag{7}$$

where Eqs. (A.11) – (A.14) are hydrogen production constraints related to the number of production plants and hydrogen production rate, and Eqs. (A.15) – (A.19) represent hydrogen storage constraints related to the number of storage units and hydrogen flowrate to the storage units. Eqs. (A.20) – (A.25) describe the pipelines transmission constraints for H_2, onshore and offshore CO_2. Eqs. (A.26) and (A.27) represent the H_2 and CO_2 energy supply–demand balance constraints, respectively. The renewable generation constraints are denoted by Eqs. (A.28) – (A.32). The H_2 imports constraints, fuel consumption constraints, CO_2 storage and emission constraints are expressed by Eqs. (A.33) – (A.39), respectively.

In the static robust optimisation (7), all investment and operational decisions are made "here-and-now" before the demand uncertainty realisations. The fundamental principle of robust decision-making is to find a solution that is feasible for all possible situations within the uncertainty range. In our model, the demand uncertainty exists in hydrogen balance constraint (A.26), which must remain feasible for all the uncertainty realisations within the polyhedral uncertainty set.

By considering the worst-case uncertainty realisation, Eq. (A.26) can be rewritten as Eq. (8):

$$\begin{aligned} \max_{D_{gtch} \in \mathcal{D}^{\mathrm{PUS}}_{gtch}} D_{gtch} \le {} & \sum_{p \in \mathcal{P}} Pr_{pgtch} + \sum_{g' \in \mathcal{N}^{\mathrm{pipe}}_{g'g}} Q_{g'gtch} + \sum_{s \in GS_{gs}} Q^{\mathrm{R}}_{sgtch} \\ & + I_{gtch} - \sum_{g' \in \mathcal{N}^{\mathrm{pipe}}_{gg'}} Q_{gg'tch} - \sum_{s \in GS_{gs}} Q^{\mathrm{I}}_{gstch}, \quad \forall g, t, c, h \end{aligned} \tag{8}$$

where $\mathcal{D}^{\mathrm{PUS}}_{gtch}$ is the uncertainty set for demand D_{gtch} at hour h, which is a subset of $\mathcal{D}^{\mathrm{PUS}}_{gtc}$. Note that the worst-case demand value is taken into account in (A.26) in order to guarantee the generation can satisfy the hydrogen demand under any uncertain situation at any time h.

Using the definition of polyhedral uncertainty sets shown in Eq. (5), i.e, $D_{gtch} = \boldsymbol{\mu}_{gtch} + \mathbf{P}_{gtch} \cdot \boldsymbol{\xi}_{gtc}$ which is hth row of $\mathbf{D}_{gtc}$, the left-hand side of Eq. (8) is written as Eq. (9):

$$\begin{aligned} & \max_{D_{gtch} \in \mathcal{D}^{\mathrm{PUS}}_{gtch}} D_{gtch} \\ & = \boldsymbol{\mu}_{gtch} + \max_{\mathbf{z}^-_{gtc}, \mathbf{z}^+_{gtc} \in \mathcal{Z}_{gtc}} \mathbf{P}_{gtch}\left(\underline{\boldsymbol{\xi}}_{gtc} \circ \mathbf{z}^-_{gtc} + \overline{\boldsymbol{\xi}}_{gtc} \circ \mathbf{z}^+_{gtc}\right) \\ & = \boldsymbol{\mu}_{gtch} + \max_{\mathbf{z}^-_{gtc}, \mathbf{z}^+_{gtc} \in \mathcal{Z}_{gtc}} \left\{ (\mathbf{P}_{gtch} \circ \underline{\boldsymbol{\xi}}^{\top}_{gtc}) \mathbf{z}^-_{gtc} + (\mathbf{P}_{gtch} \circ \overline{\boldsymbol{\xi}}^{\top}_{gtc}) \mathbf{z}^+_{gtc} \right\} \end{aligned} \tag{9}$$

where $\boldsymbol{\mu}_{gtch} \in \mathbb{R}$ and $\mathbf{P}_{gtch} \in \mathbb{R}^{1\times 24}$ are the hth row elements of $\boldsymbol{\mu}_{gtc}$ and $\mathbf{P}_{gtc}$ defined in Section 3.2, respectively, and $\mathcal{Z}_{gtc} = \left\{\mathbf{z}^-_{gtc}, \mathbf{z}^+_{gtc} \mid \mathbf{0} \le \mathbf{z}^-_{gtc}, \mathbf{z}^+_{gtc} \le \mathbf{1}, \mathbf{1}^{\top}(\mathbf{z}^-_{gtc} + \mathbf{z}^+_{gtc}) \le \boldsymbol{\Phi}_{gtc}, \mathbf{z}^-_{gtc} + \mathbf{z}^+_{gtc} \le \mathbf{1}\right\}$. The optimisation problem in right-hand side of (9) is then converted to Eq. (10):

$$\begin{aligned} & \max_{\mathbf{z}^-_{gtc}, \mathbf{z}^+_{gtc}} \left\{ (\mathbf{P}_{gtch} \circ \underline{\boldsymbol{\xi}}^{\top}_{gtc}) \mathbf{z}^-_{gtc} + (\mathbf{P}_{gtch} \circ \overline{\boldsymbol{\xi}}^{\top}_{gtc}) \mathbf{z}^+_{gtc} \right\} \\ & \text{s.t. } \mathbf{z}^-_{gtc} \le \mathbf{1}, \quad \mathbf{z}^+_{gtc} \le \mathbf{1}, \quad \mathbf{z}^-_{gtc} + \mathbf{z}^+_{gtc} \le \mathbf{1} \\ & \qquad \mathbf{1}^{\top}(\mathbf{z}^-_{gtc} + \mathbf{z}^+_{gtc}) \le \boldsymbol{\Phi}_{gtc} \\ & \qquad \mathbf{z}^-_{gtc}, \mathbf{z}^+_{gtc} \ge \mathbf{0} \quad \forall g, t, c, h, \end{aligned} \tag{10}$$

which is equivalent to the following minimisation problem (11) by using the duality theory, since it is a linear program of decision variables $\mathbf{z}^-_{gtc}, \mathbf{z}^+_{gtc}$. By defining $\underline{\mathbf{Q}}_{gtch} = \mathbf{P}_{gtch} \circ \underline{\boldsymbol{\xi}}^{\top}_{gtc} \in \mathbb{R}^{1\times 24}$ and $\overline{\mathbf{Q}}_{gtch} = \mathbf{P}_{gtch} \circ \overline{\boldsymbol{\xi}}^{\top}_{gtc} \in \mathbb{R}^{1\times 24}$, respectively, we have the dual problem:

$$\begin{aligned} \min_{\lambda^-_{gtch}, \lambda^+_{gtch}, \boldsymbol{\gamma}_{gtch}, \varrho_{gtch}} \quad & \mathbf{1}^{\top}(\lambda^-_{gtch} + \lambda^+_{gtch} + \boldsymbol{\gamma}_{gtch}) + \varrho_{gtch} \boldsymbol{\Phi}_{gtc} \\ \text{s.t.} \quad & \lambda^-_{gtch} + \boldsymbol{\gamma}_{gtch} + \varrho_{gtch} \mathbf{1} \ge \underline{\mathbf{Q}}^{\top}_{gtch} \\ & \lambda^+_{gtch} + \boldsymbol{\gamma}_{gtch} + \varrho_{gtch} \mathbf{1} \ge \overline{\mathbf{Q}}^{\top}_{gtch} \\ & \lambda^-_{gtch}, \lambda^+_{gtch}, \boldsymbol{\gamma}_{gtch} \ge \mathbf{0}, \quad \varrho_{gtch} \ge 0, \quad \forall g, t, c, h \end{aligned} \tag{11}$$

where $\lambda^-_{gtch}, \lambda^+_{gtch}, \boldsymbol{\gamma}_{gtch} \in \mathbb{R}^{24}$ are the dual variables (Lagrange multipliers) to the first three constraints in (10), respectively, and $\varrho_{gtch} \in \mathbb{R}$ is the dual variable to the fourth constraint.

Combining the minimisation problem in (11) with Eq. (8), we can get Eq. (12)

$$\begin{aligned} & \boldsymbol{\mu}_{gtch} + \mathbf{1}^{\top}(\lambda^-_{gtch} + \lambda^+_{gtch} + \boldsymbol{\gamma}_{gtch}) + \varrho_{gtch} \boldsymbol{\Phi}_{gtc} \\ \le {} & \sum_{p \in \mathcal{P}} Pr_{pgtch} + \sum_{g' \in \mathcal{N}^{\mathrm{pipe}}_{g'g}} Q_{g'gtch} + \sum_{s \in GS_{gs}} Q^{\mathrm{R}}_{sgtch} \\ & + I_{gtch} - \sum_{g' \in \mathcal{N}^{\mathrm{pipe}}_{gg'}} Q_{gg'tch} - \sum_{s \in GS_{gs}} Q^{\mathrm{I}}_{gstch}, \quad \forall g, t, c, h. \end{aligned} \tag{12}$$

Note that the minimisation term of (11) is omitted after substituting it to the left-hand side of Eq. (8), thus forming Eq. (12), because it is sufficient that the constraint holds for at least one $(\lambda^-_{gtch}, \lambda^+_{gtch}, \boldsymbol{\gamma}_{gtch}, \varrho_{gtch})$ satisfying the constraints in (11). The logic behind it can be demonstrated from the compact case: $\inf_{A^{\top}p \le B} b^{\top}p \le 0 \Leftrightarrow b^{\top}p \le 0\ \exists p$ s.t. $A^{\top}p \le B$ where p is the decision variable.

After obtaining the above constraint, we formulate a data-driven robust counterpart for optimisation problem (7), as shown below.

Problem 1 (*SRO*).

$$\begin{aligned}
&\min\ TC\\
&\text{s.t. (6), (12), (A.1) – (A.25), (A.27) – (A.39)}\\
&\quad \lambda^{-}_{gtch} + \boldsymbol{\gamma}_{gtch} + \varrho_{gtch}\mathbf{1} \geq \underline{\mathbf{Q}}^{\top}_{gtch}\\
&\quad \lambda^{+}_{gtch} + \boldsymbol{\gamma}_{gtch} + \varrho_{gtch}\mathbf{1} \geq \overline{\mathbf{Q}}^{\top}_{gtch}\\
&\quad \lambda^{-}_{gtch}, \lambda^{+}_{gtch}, \boldsymbol{\gamma}_{gtch} \geq \mathbf{0},\\
&\quad \varrho_{gtch} \geq 0, \quad \forall g \in \mathcal{G}, t \in \mathcal{T}, c \in \mathcal{C}, h \in \mathcal{H},
\end{aligned} \tag{13}$$

which is an MILP problem, and can be solved using the off-the-shelf method like branch-and-cut method implemented in solver Gurobi [60].

Compared to the deterministic optimisation, the static robust optimisation further introduces more decision variables, the size of which is summarised later in Table 2 of Section 6. To alleviate its computational complexity, we adopt a two-step hierarchical procedure introduced in [10,24] to handle it by relaxing the part of integer variables. The first step consists of the solution to the original problem without the pipelines investments, and then the production and storage investment decisions $NP_{pgt}, IP_{pgt}, NS_{sgt}, IS_{sgt}$ obtained from the first step are fixed and transmitted to the second step to solve the full model in (13) so as to calculate the remaining integer and continuous decision variables. As verified in [24], the two-step procedure can guarantee a near-optimal solution while significantly reducing the computing time.

The solution of Problem 1 provides a robust strategy guaranteeing that the energy balance constraint is feasible for all uncertainty realisations in the uncertainty set. The static robust framework suits the case that all the decisions need to be made before all the uncertainty realisations. But some of the operational decisions in the hydrogen planning problem can be made after the uncertainty realisations. In addition, the solution to (13) is overly conservative because it is unlikely that all the worst-case scenarios for each hour t happen at the same time. Therefore, we introduce the data-driven adaptive robust optimisation in the next section to overcome these issues.

5. Data-driven adaptive robust optimisation

The data-driven adaptive robust optimisation (ARO) allows for the dynamic adjustment of operational decisions to hedge against the uncertain risk, and provides a more flexible framework to control the level of conservativeness of the robust solution. In ARO, the decisions are made in a "wait-and-see" manner with the uncertainty feedback. Hence, ARO can better model the sequential decision-making process and avoid the overly-conservative issues of SRO. It has been widely used in the energy system such as the power system expansion planning and demand response of multi-energy hubs [31,61] and its effectiveness has been demonstrated as well. In addition, ARO only requires the range of the uncertain data, which avoids the difficult capture of probability distributions required for scenario-based approaches in stochastic programming. Therefore, we are motivated to develop an ARO framework for the large-scale hydrogen infrastructure planning problem.

In general, the investment decisions are made on the yearly scale before the demand uncertainty realisations, while the operational decisions are made on the hourly level and can be adjusted after the uncertainty realisations, which fits well into the sequential decision-making process and can lead to a more adaptive and less conservative solution. To this end, we reformulate the optimisation problem in (7) to adapt to the ARO framework, which is shown in the following Problem 2. We define a composite vector ($\mathbf{x}$) related to investment decisions of production and storage technologies and transmission pipelines

$$\begin{aligned}
\mathbf{x} = \big(&IP_{pgt}, NP_{pgt}, IS_{sgt}, NS_{sgt}, AY_{gg't}, AY^{\mathrm{S}}_{g,sa,t}, \overline{AY}_{gg't}, \overline{\overline{AY}}_{grt},\\
&Y_{gg't}, Y^{\mathrm{S}}_{g,sa,t}, \overline{Y}_{gg't}, \overline{\overline{Y}}_{grt}, IR_{egt}, NR_{egt}, RI_{rt}, \forall p, s, e, g, t, r\big) \in \mathcal{X}
\end{aligned}$$

where $\mathcal{X}$ is a set of $\mathbf{x}$ satisfying local constraints that only related to the composite variable $\mathbf{x}$. We then define a composite continuous vector ($\mathbf{y}$) related to the operational decisions

$$\begin{aligned}
\mathbf{y} = \big(&CL_{gtch}, I_{gtch}, Pr_{pgtch}, \widehat{Pr}_{egtch}, Q_{gg'tch}, Q^{\mathrm{I}}_{gstch}, Q^{\mathrm{R}}_{sgtch},\\
&\overline{Q}_{gg'tch}, \overline{\overline{Q}}_{grtch}, St_{sgtch}, \forall p, s, e, g, t, r, c, h\big) \in \Omega(\mathbf{x}, \mathbf{D})
\end{aligned}$$

where $\Omega(\mathbf{x}, \mathbf{D})$ is a set of $\mathbf{y}$ satisfying its local constraints and linking constraints that related to variables $\mathbf{x}, \mathbf{y}$ and uncertain demand $\mathbf{D} = (D_{gtch}; g \in \mathcal{G}, t \in \mathcal{T}, c \in \mathcal{C}, h \in \mathcal{H})$. The variable $\mathbf{y}$ is composed of the continuous decision variables at the lower-level of the ARO model. They represent the renewable power curtailment, the hydrogen flowrate, hydrogen production rate, CO_2 flowrate and hydrogen storage state at each hour t, respectively. A detailed description of decision variables related to $\mathbf{x}$ and $\mathbf{y}$ composite vectors can be found in the Appendix and Nomenclature.

Problem 2 (*ARO*).

$$\begin{aligned}
&\min_{\mathbf{x}\in\mathcal{X}} \Big\{ PCC + SCC + PLCC + POC1 + SOC1 + PLOC + ReC\\
&\quad + \max_{\mathbf{D}\in\mathcal{D}^{\mathrm{PUS}}} \min_{\mathbf{y}\in\Omega(\mathbf{x},\mathbf{D})} \big\{ POC2 + SOC2 + CEC + IIC + FC \big\} \Big\}\\
&\text{s.t. (A.1) – (A.10)}, \quad \mathcal{D}^{\mathrm{PUS}} = \Big\{ D^{\mathrm{PUS}}_{gtc}; g \in \mathcal{G}, t \in \mathcal{T}, c \in \mathcal{C} \Big\}\\
&\quad \mathcal{X} = \Big\{ \mathbf{x} \,\Big|\, \text{(A.11), (A.15), (A.20) – (A.22), (A.31), (A.32), (A.37)} \Big\}\\
&\quad \Omega(\mathbf{x}, \mathbf{D}) = \Big\{ \mathbf{y} \,\Big|\, \text{(A.12) – (A.14), (A.16) – (A.19), (A.23) – (A.30), (A.33) – (A.36), (A.38), (A.39)} \Big\}
\end{aligned} \tag{14}$$

where cost $POC1/SOC1$ denotes the first term of right-hand side of Eq. (A.5)/Eq. (A.6) that related to integer variables (investment decisions), while cost $POC2/SOC2$ denotes its second term that related to continuous variables (operational decisions), accordingly, such that $POC = POC1 + POC2$ and $SOC = SOC1 + SOC2$.

Problem 2 is a tri-level optimisation problem. The first minimisation problem is called the upper level problem which consists of the annualised investment cost and the worst-case operating cost. The investment decisions are subject to the constraints shown in $\mathcal{X}$. The middle-level problem, i.e., the second maximisation problem identifies the uncertainty realisations under worst-case scenario. The feasible region for this level is defined in Section 3.2. With the upper and middle-level decision variables, the operating cost is minimised in the lower-level subject to the constraints shown in $\Omega(\mathbf{x}, \mathbf{D})$. The goal of this min–max–min formulation is to protect the system against the worst-case realisation of uncertain demand within the polyhedral uncertainty set, i.e., to confer resilience upon the system for all potential instances encompassed by the uncertainty set.

It is worth noting that in the above robust framework while the clusters are introduced for reducing the model size, the proposed model is different from the robust-stochastic modelling framework in the literature [55,62,63] which combines the stochastic programming (SP) with the (A)RO for different uncertain parameters in the optimisation problem. This work does not involve SP at all for demand uncertainty. Considering the computational complexity of the underlying multi-scale problem, the above (A)RO-based framework is proposed to systematically generate representative days with the related uncertainty sets using data-driven methods. The established uncertainty sets consider the data correlations between hours, which is different from the robust-stochastic models in the literature.

5.1. Hybrid decomposition for ARO problem

The above formulated tri-level ARO problem cannot be solved directly using off-the-shelf optimisation solvers. To this end, we employ the widely adopted column-and-constraint generation (CCG) algorithm [64,65] to decompose Problem 2 into a "min" master problem and a "max–min" sub-problem.

The CCG master problem is a relaxation of original Problem 2, which is an MILP problem as well. A set of operating constraints related to the demand under worst-case scenario obtained from the sub-problem at iteration $K-1$ are added into the master problem at iteration K. We can obtain tentative investment decisions after solving the master problem, and then transmit these decisions to the CCG sub-problem to calculate the worst-case scenario. The master problem provides a lower bound while the sub-problem determines the upper bound. The whole process continues until the gap between the upper and lower bounds iteratively reaches an optimality tolerance.

5.1.1. CCG master problem

The CCG master problem can be formulated as follows:

$$\min_{\mathbf{x},\eta,\mathbf{y}^k} \left\{ PCC + SCC + PLCC + POC1 + SOC1 + PLOC + ReC + \eta \right\}$$

s.t. (A.1) – (A.11), (A.15), (A.20) – (A.22), (A.31), (A.32), (A.37)

$$\eta \geq POC2^k + SOC2^k + CEC^k + IIC^k + FC^k,$$

$$Pr^k_{pgtch} \leq cap^{\max}_p \cdot NP_{pgt} \quad \forall p,g$$

$$Pr^k_{pgtch} - Pr^k_{pgtc,h-1} \leq RU_p \cdot cap^{\mathrm{P}}_p \cdot NP_{pgt} \quad \forall p,g$$

$$Pr^k_{pgtc,h-1} - Pr^k_{pgtch} \leq RD_p \cdot cap^{\mathrm{P}}_p \cdot NP_{pgt} \quad \forall p,g$$

$$cap^{\min}_s \cdot NS_{sgt} \leq St^k_{sgtch} \leq cap^{\max}_s \cdot NS_{sgt},$$

$$St^k_{sgtch} = St^k_{sgtc,h-1} + Q^{\mathrm{I},k}_{gstch} - Q^{\mathrm{R},k}_{sgtch}, \quad \forall \{s,g\} \in GS_{gs}$$

$$St^k_{sgtc,24} = St^{\mathrm{init}}_{sgtc,1} \quad \forall \{s,g\} \in GS_{gs}$$

$$Q^{\mathrm{I},k}_{gstch} \leq Q^{\mathrm{Imax}}_s \cdot NS_{sgt} \quad \forall \{s,g\} \in GS_{gs}$$

$$Q^{\mathrm{R},k}_{gstch} \leq Q^{\mathrm{Rmax}}_s \cdot NS_{sgt} \quad \forall \{s,g\} \in GS_{gs}$$

$$Q^k_{gg'tch} \leq q^{\mathrm{Hmax}} \cdot AY_{gg't} \quad \forall g,g' \in \mathcal{N}^{\mathrm{pipe}}_{gg'}$$

$$\overline{Q}^k_{gg'tch} \leq q^{\mathrm{Cmax}} \cdot \overline{AY}_{gg't} \quad \forall g,g' \in \mathcal{N}_{gg'}$$

$$\overline{\overline{Q}}^k_{grtch} \leq q^{\mathrm{Cmax}} \cdot \overline{\overline{AY}}_{grt} \quad \forall \{g,r\} \in GR_{gr}$$

$$\sum_{p\in\mathcal{P}} Pr^k_{pgtch} + \sum_{g'\in\mathcal{N}^{pipe}_{g'g}} Q^k_{g'gtch} + \sum_{s\in GS_{gs}} Q^{\mathrm{R},k}_{sgtch} + I^k_{gtch} \geq \sum_{g'\in\mathcal{N}^{\mathrm{pipe}}_{gg'}} Q^k_{gg'tch} + \sum_{s\in GS_{gs}} Q^{\mathrm{I},k}_{gstch} + D^k_{gtch}, \quad \forall g,t,c,h$$

$$\sum_{g'\in\mathcal{N}_{g'g}} \overline{Q}^k_{g'gtch} + \sum_{p\in\mathcal{P}} y^c_{pt} Pr^k_{pgtch} = \sum_{g'\in\mathcal{N}_{gg'}} \overline{Q}^k_{gg'tch} + \sum_{r\in GR_{gr}} \overline{\overline{Q}}^k_{grtch}, \quad \forall g,t,c,h$$

$$Pr^k_{pgtch} = \eta_t \Big(\sum_{e\in\mathcal{E}} \widehat{Pr}^k_{egtch} - CL^k_{gtch} \Big), \quad \forall p \in \{WE\}$$

$$\sum_{g\in\mathcal{G}} \sum_{t\in\mathcal{T}} CL^k_{gtch} \leq cl \cdot \sum_{e\in\mathcal{E}} \sum_{g\in\mathcal{G}} \sum_{t\in\mathcal{T}} \widehat{Pr}^k_{egtch}, \quad \forall c,h$$

$$\widehat{Pr}^k_{egtch} = AV_{egch} \cdot NR_{egt}, \quad \forall e \in \mathcal{E}, g,t,c,h$$

$$I^k_{gtch} \leq I^{\mathrm{Up}}_{gtc}, \quad \forall g \in Gimp_g, t,c,h$$

$$\sum_{p\in\{BGCCS\}} \sum_{c\in C} \sum_{h\in\mathcal{H}} WF_c \cdot \frac{Pr^k_{pgtch}}{\eta_{pt}} \leq ba_{gt}, \quad \forall g,t$$

$$RI_{rt} = RI_{r,t-1} + \tau \sum_{g\in GR_{gr}} \sum_{c\in C} \sum_{h\in\mathcal{H}} WF_c \cdot \overline{\overline{Q}}^k_{grtch}, \quad \forall r \in \mathcal{R}, t$$

$$\sum_{p\in\mathcal{P}} \sum_{g\in\mathcal{G}} \sum_{c\in C} \sum_{h\in\mathcal{H}} WF_c \cdot y^e_{pt} \cdot Pr^k_{pgtch} \leq et_t$$

$$k = 1, \ldots, K, \quad \forall t,c,h \tag{15}$$

where η in the objective function is the approximation of primal objective function at lower level, i.e., the approximation of operating cost in Problem 2. The decisions in (15) include the upper level decision variables $\mathbf{x}$ and also the additional variable $\mathbf{y}^k$ corresponding to $\mathbf{y}$ at each iteration $k = 1, \ldots, K$. Except the local constraints for $\mathbf{x}$, a set of primal cuts constructed by $\mathbf{y}^k$ is added to the master problem. Note that the obtained values of η and $\mathbf{y}^k$ after solving master problem (15) will not be transmitted to the CCG sub-problem, and only the value of $\mathbf{x}$ is needed.

We can observe that the CCG master problem is a complex MILP problem and its scale will grow as the outer iteration step K proceeds, which could make it more and more computationally intensive. Hence, we still use the two-step hierarchical procedure introduced in Section 4 to solve the problem (15) to decompose it into two optimisation problems that can be solved relatively quickly. The solution process is similar to that in Section 4.

5.1.2. CCG subproblem

The CCG sub-problem is a bi-level "max–min" optimisation problem parameterised in terms of the given investment decisions $\mathbf{x}^K$ at iteration K, which aims to extract the worst-case realisations of uncertain demand. It can be formulated as follows where all the decision variables are continuous and constraints are linear:

$$\Xi(\mathbf{x}^K) = \max_{\mathbf{D}\in\mathcal{D}^{\mathrm{PUS}}} \min_{\mathbf{y}} \left\{ POC2 + SOC2 + CEC + IIC + FC \right\}$$

s.t. (A.5) – (A.8), (A.10), (A.17), (A.26) – (A.29), (A.33) – (A.35), (A.38), (A.39),

$$Pr_{pgtch} \leq cap^{\max}_p \cdot NP^K_{pgt}, \quad \forall p,g,t,c,h$$

$$Pr_{pgtch} - Pr_{pgtc,h-1} \leq RU_p \cdot cap^{\mathrm{P}}_p \cdot NP^K_{pgt}, \quad \forall p,g,t,c,h$$

$$Pr_{pgtc,h-1} - Pr_{pgtch} \leq RD_p \cdot cap^{\mathrm{P}}_p \cdot NP^K_{pgt}, \quad \forall p,g,t,c,h$$

$$cap^{\min}_s \cdot NS^K_{sgt} \leq St_{sgtch} \leq cap^{\max}_s \cdot NS^K_{sgt}, \quad \forall s,g,t,c,h$$

$$Q^{\mathrm{I}}_{gstch} \leq Q^{\mathrm{Imax}}_s \cdot NS^K_{sgt} \quad \forall \{s,g\} \in GS_{gs}, t,c,h$$

$$Q^{\mathrm{R}}_{gstch} \leq Q^{\mathrm{Rmax}}_s \cdot NS^K_{sgt} \quad \forall \{s,g\} \in GS_{gs}, t,c,h$$

$$Q_{gg'tch} \leq q^{\mathrm{Hmax}} \cdot AY^K_{gg't} \quad \forall g,g' \in \mathcal{N}^{\mathrm{pipe}}_{gg'}, t,c,h$$

$$\overline{Q}_{gg'tch} \leq q^{\mathrm{Cmax}} \cdot \overline{AY}^K_{gg't} \quad \forall g,g' \in \mathcal{N}_{gg'}, t,c,h$$

$$\overline{\overline{Q}}_{grtch} \leq q^{\mathrm{Cmax}} \cdot \overline{\overline{AY}}^K_{grt} \quad \forall \{g,r\} \in GR_{gr}, t,c,h$$

$$\widehat{Pr}_{egtch} = AV_{egch} \cdot NR^K_{egt} \quad \forall e \in \mathcal{E}, g,t,c,h$$

$$RI^K_{rt} = RI^K_{r,t-1} + \tau \sum_{g\in GR_{gr}} \sum_{c\in C} \sum_{h\in\mathcal{H}} WF_c \cdot \overline{\overline{Q}}_{grtch}, \quad \forall r \in \mathcal{R}, t. \tag{16}$$

Note that the above bi-level optimisation problem is NP-hard and cannot be directly solved using the off-the-shelf solvers. The common way to handle it is converting the "max–min" problem into a single-level "max" optimisation problem with bilinear terms by using the dualisation theory [30,53]. Then a big-M method is introduced for the linearisation with additional binary variables. Nevertheless, the choice of appropriate big-M values is a challenging task [66], and the introduction of additional variables results in the solution of a combinatorial program for the CCG sub-problem.

Therefore, we propose to use the Block coordinate descent (BCD) method to efficiently solve the sub-problem (16), which does not need the introduction of dualisation and big-M values [31]. BCD decomposes the "max–min" problem into two linear program problems: the lower level problem determining the operating decisions with fixed demand, and the middle level problem determining the uncertain demand with fixed decisions, and these two LP problems are solved iteratively until the BCD method convergence.

The BCD lower-level problem is expressed as follows:

$$\Phi_{lo} = \min_{\mathbf{y}} \left\{ POC2 + SOC2 + CEC + IIC + FC \right\}$$

s.t. All constraints in (16)

$$D_{gtch} = D^{v}_{gtch} \quad : \mu_{gtch}, \quad \forall g,t,c,h \tag{17}$$

where $\mu_{gtch}, \forall g,t,c,h$ are the dual variables, which quantify the sensitivity of the objective function with respect to variations in demand at iteration step v of BCD method. The dual variables will be employed

in the BCD middle-level problem to get the worst-case uncertainty realisation.

The BCD middle-level problem is formulated below:

$$\max_{\mathbf{D}\in\mathcal{D}^{\mathrm{PUS}}} \ {\Phi_{lo}}^{v-1} + \sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{c\in\mathcal{C}}\sum_{h\in\mathcal{H}} \mu_{gtch}^{v-1}\cdot(D_{gtch} - D_{gtch}^{v-1}) \tag{18}$$

where the demand is subject to the polyhedral uncertainty set, μ_{gtch}^{v-1} and ${\Phi_{lo}}^{v-1}$ are obtained from the lower-level at iteration $v-1$. This middle-level problem is built upon the first-order Taylor series approximation of the operating cost Φ_{lo} around the uncertainty realisations at the previous iteration.

5.1.3. Summary of ARO solution algorithm

The above iterative procedure forms our proposed algorithm called *Enhanced column-and-constraint generation (ECCG) algorithm*, which is given below. It involves one outer loop related to the traditional CCG algorithm and two inner loops associated to the two-step hierarchical method for solving CCG master problem (CCG-MP) and BCD method for solving CCG sub-problem (CCG-SP), as shown in flowchart Fig. 4. Initialise the upper bound $UB = +\infty$ and lower bound $LB = -\infty$, and set tolerance $\varepsilon_1 > 0$ for the outer loop and $\varepsilon_2 > 0$ for the inner BCD method.

(1) Initialisation of the outer loop: Set the iteration counter K to 1 and the maximum iteration step $K^{\max}$; Set the initial investment decisions $\mathbf{x}^K$.
(2) Initialisation of the inner loop: Set the iteration counter of inner loop about the BCD method v to 1, and initialise demand values $\mathbf{D}^v$; Set the operational cost for the inner loop C^{IL} to $+\infty$.
(3) Solve the BCD lower-level problem (17) with given $\mathbf{x}^K$ and $\mathbf{D}^v$, which provides ${\Phi_{lo}}^v$ and μ_{gtch}^v.
(4) Check the convergence of inner BCD method: If $|C^{IL} - {\Phi_{lo}}^v|/|C^{IL}| \le \varepsilon_2$, end the BCD iteration and set $\mathbf{D}^K \leftarrow \mathbf{D}^v$, and then go to Step (6); otherwise, continue the BCD iteration and set $C^{IL} \leftarrow {\Phi_{lo}}^v$, update the BCD iteration counter $v \leftarrow v+1$.
(5) Solve the BCD middle-level problem in (18) with given ${\Phi_{lo}}^{v-1}$ and μ_{gtch}^{v-1}, which provides $\mathbf{D}^v$. Then go to Step (3).
(6) Update the outer loop iteration counter $K \leftarrow K+1$.
(7) Solve the CCG master problem in (15) with given $\mathbf{D}^v$ by using the two-step hierarchical procedure, which provides the investment decisions $\mathbf{x}^K$; Assign the objective value of the CCG master problem to LB, and calculate $UB = LB - \eta^K + C^{IL}$.
(8) Check the convergence of outer loop: If $|UB - LB|/|LB| \le \varepsilon_1$ and $K \le K^{\max}$, terminate the whole algorithm; otherwise, go to Step (2).

Fig. 4 showcases how the proposed ECCG algorithm works. Firstly, instead of solving a monolithic large-scale MILP directly, we use a two-step hierarchical procedure to decompose the master problem to two relatively small-scale MILP problems. Step 1 allows for the solution without the pipeline network constraints, which decreases the model size significantly, and then with the obtained production and storage investment decisions from Step 1 the full model with pipeline constraints is calculated at Step 2. This hierarchical procedure allows for significantly faster computations, as will be demonstrated later on in Table 2. Secondly, the BCD method decomposes the "max–min" subproblem to two computationally straightforward LPs rather than solving an underlying MILP problem resulted from the big-M formulation, which further improves the computational efficiency. Overall, compared to CCG method, ECCG is more convenient to implement, particularly for multi-scale complex models. It offers improved scalability and reduces computational complexity.

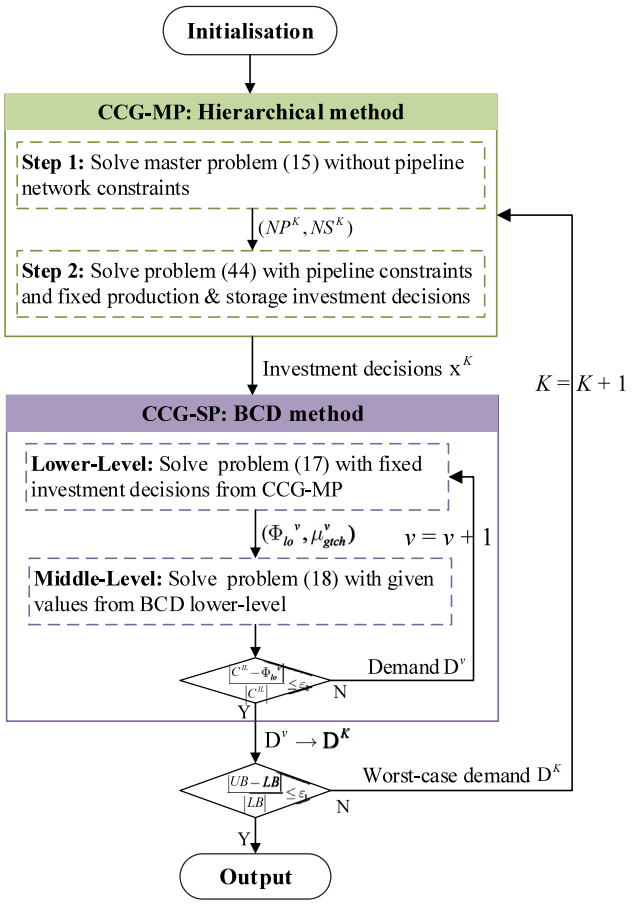


Fig. 4. The flowchart of the proposed ECCG algorithm.

6. Computational results

We conduct the numerical simulations in this section to evaluate the performance of the proposed framework for the hydrogen infrastructure planning in GB shown in Fig. 2 over the 5-year steps 2035–2050. In order to derive the uncertainty sets for hydrogen demand, historical regional hourly gas consumption data are collected from the GB gas distribution companies for several years (2015–2018) [6]. These historical data account for industrial, domestic and commercial loads, which were further dealt with to segment heat-related hydrogen demands within each region for heat decarbonisation in GB. We use the k-medoids clustering method introduced in Section 3.1 to disaggregate each year's hydrogen demand into several clusters. The peak demand day is treated as one cluster for system's security reasons. We establish the polyhedral uncertainty sets for other remaining clusters by applying the method shown in Section 3.2. All simulations are executed in GAMS Studio 1.13.4 on a PC with Intel® Core™ i9-10980XE CPU @ 3.00 GHz and 128 GB RAM, and solved by Gurobi 9.5.1 with optimality gap 5%. Model parameters kept the same as our previous paper [24] and can be found in its supplementary material. The lifetimes of equipment for production, storage and transmission are set as 30–40 years, which is taken from the report of UK BEIS [67]. The planning interval is less than the lifetimes of technologies and pipelines, and hence the terminal residual values of equipment are not deducted from the investment cost. The relative tolerance for the BCD method is 1×10^{-8}. The optimality tolerance of outer hybrid decomposition algorithm is set to be 0.1%. The solution results of the multi-scale model find the optimal yearly investment decisions and hourly operational decisions for the hydrogen infrastructure network.

Table 1
Cost comparison between deterministic and ARO scenarios under 4 clusters and $\Phi_{gtc} = 3$.

Scenarios	Total cost (£b)	Running time (min)
Deterministic	63.08	9.73
SRO	114.54	9.82

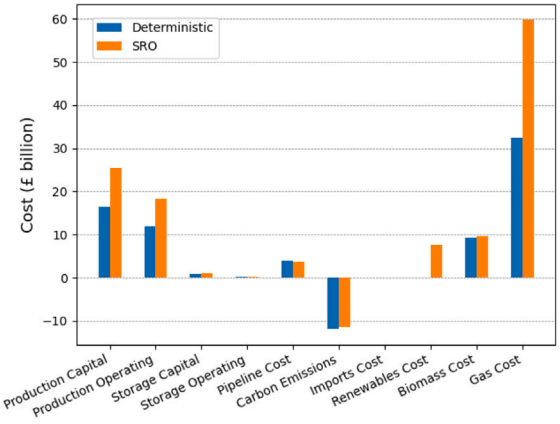


Fig. 5. Cost comparison between deterministic model and SRO model under 4 clusters and $\Phi_{gtc} = 3$.

6.1. Comparison between deterministic and SRO models

We compare the model performances with and without uncertainty with 4 clusters in this section. The deterministic model means that the uncertainty budget value Φ_{gtc} of polyhedral uncertainty set introduced in (5) is 0, and the hydrogen demand in each cluster is equal to the average values, which neglects the daily variability of demand. In SRO, we set the value of budget Φ_{gtc} as 3 because the resulting data-driven uncertainty set can cover the majority of uncertainty data. This leads to more hydrogen demand because the SRO needs to guarantee that the solution is feasible for all uncertainty realisations within an uncertainty set. The SRO model is solved as an MILP model by using the two-step hierarchical method introduced in Section 4, which includes 544 discrete variables, 472,800 continuous variables and 419,832 equations.

As we can see from Table 1, the total system cost increases by around 81.5% using the SRO approach compared with the deterministic scenario because of increased demand. The total CPU running time of these two scenarios is similar under 4 clusters. The spatially-explicit pattern of hydrogen demand considers trade-offs between production, storage, transmission and emissions costs. Fig. 5 shows the detailed cost with around 50% increase in production capital and operating cost and 85% increase in gas cost by using robust approach compared with the deterministic model. The biomass cost for both of them is similar because of limited biomass availability, while most of hydrogen is produced by consuming gas. Neither of them involves hydrogen imports as seen from Fig. 5. Note that SRO model leads to the investment and use of renewable energies as well to produce more hydrogen by WE technology. Fig. 6 compares the total production capacity of different technologies in different years. We can observe that ATRCCS technology is more widely used and favourable compared to SMRCCS technology. This is because even though the capital cost of ATRCCS technology is a bit higher, it demonstrates superior energy conversion efficiency and performance, potentially enabling more effective resource utilisation and lower operational cost. From 2035–2045, the total production capacity is increasing year by year due to increasing hydrogen loads, and then remains essentially the same, which can also be observed from Fig. 7 showing the temporal and spatial production for SRO. The NO region produced the most accounting for around 19% in 2035 due to the high demand, while the production percentage in other regions is almost similar in each year.

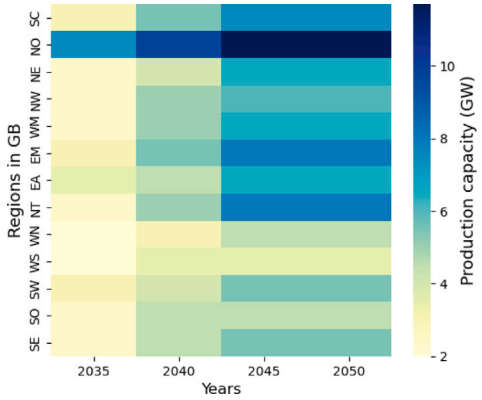


Fig. 7. Temporal and spatial heatmap of production capacity for SRO under 4 clusters and $\Phi_{gtc} = 3$.

In Fig. 8, the production capacity and hydrogen pipelines expansion maps are illustrated from 2035 to 2050 year. The total production capacity increases from 39 GW in 2035 to 63.7 GW in 2040 and further to 84.2 GW in 2050. We can observe that the hydrogen in NO region is produced not only by ATRCCS and BGCCS technologies, but also WE technology consuming renewables. Although the WE technology has higher operating cost compared with ATRCCS technology, it has relatively lower capital cost compared with ATRCCS and BGCCS technologies, and does not generate CO_2. Therefore, when the hydrogen demand is very high in some individual regions like NO region, it may be more economical to invest more WE technologies rather than more ATRCCS technologies, which may result in relatively low capital cost in this region despite potentially higher production costs, while the total cost is reduced. It is the balance of operational cost and capital cost. This case does not suit the scenario of low hydrogen demand, wherein the investment in production technology is relatively small compared to high demand scenario. However, due to the limited installation of renewable farms and limited investment number of WE technology in each region, the production of WE will not continue to increase, and perhaps at some point the operational cost will be higher than

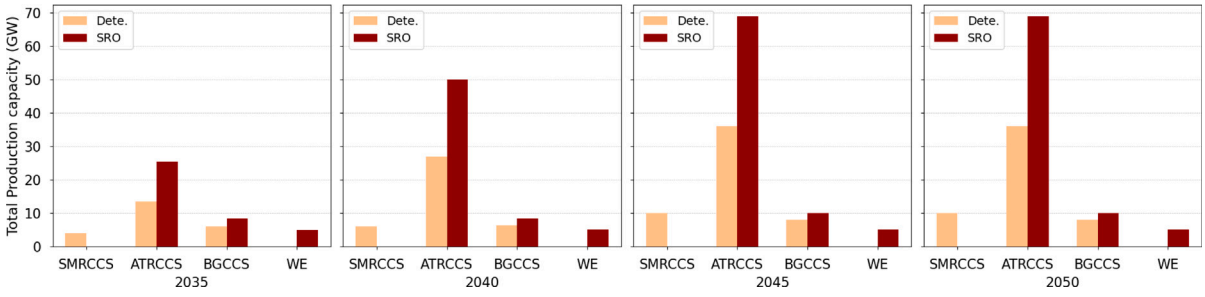
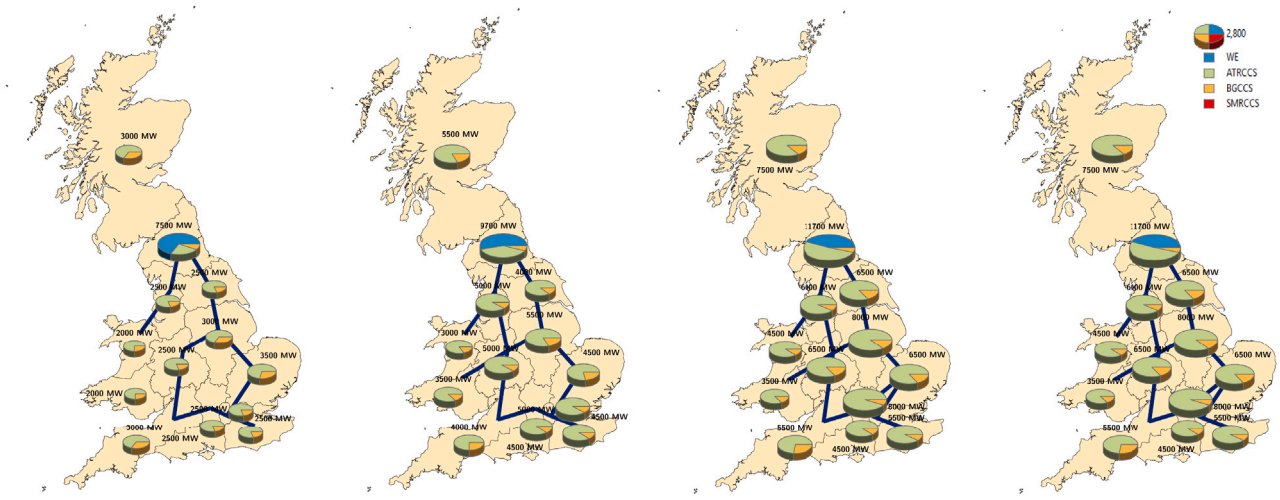


Fig. 6. The total production capacity comparison of different technologies between deterministic (Dete.) model and SRO model with $\Phi_{gtc} = 3$ under 4 clusters in different years.



(a) 2035 (b) 2040 (c) 2045 (d) 2050

Fig. 8. Production capacity and hydrogen pipelines expansion maps from 2035 to 2050 for SRO under 4 clusters and $\Phi_{gtc} = 3$.

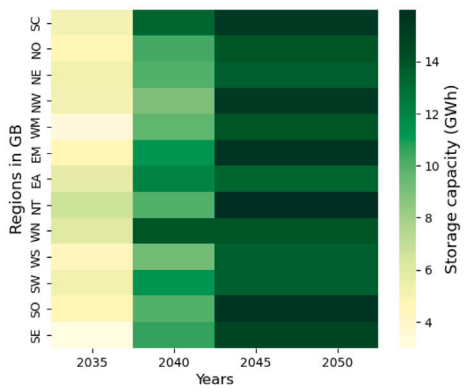


Fig. 9. Temporal and spatial heatmap of storage capacity for SRO under 4 clusters and $\Phi_{gtc} = 3$.

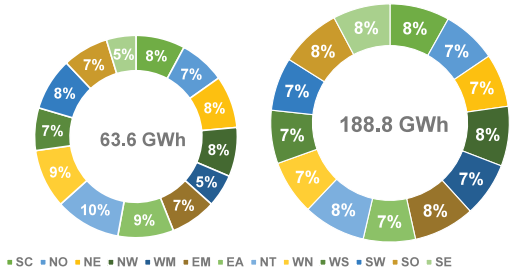


Fig. 10. The percentage of storage capacity in each region from 2035 to 2050 for SRO under 4 clusters and $\Phi_{gtc} = 3$.

the capital cost leading to an overall increase in total costs over time. That is why in Fig. 6, the production capacity of WE looks basically unchanged in different years. The transportation of hydrogen across regions is facilitated by an interconnected pipeline network. Illustrated in Fig. 8, the hydrogen transmission network undergoes incremental development from 2035, with initial installations covering only a portion of the intended network. Subsequently, the network expands progressively, establishing connections with neighbouring regions in 2045. Note that the pipelines availability in 2050 is the same as 2045 because all the pipelines have been established in 2045 and no further expansion and pipeline investment needed in 2050.

The hydrogen storage plays a significant role in "peak shaving" so as to reduce the total system cost. Fig. 9 displays the temporal and spatial storage for SRO. It can be observed that the storage capacity in each region increases annually and then remains constant, with similar annual increments across all regions. The storage technology installed in all regions of GB is MPSV with or without uncertainty, because it offers lower operating cost for storage compared with HPSV technology. The detailed storage in each region for SRO is shown in Fig. 10 with a total storage capacity 63.6 GWh in 2035 and 188.8 GWh in 2050 year, which is slightly higher than the total storage capacity 172.8 GWh in 2050 for deterministic case. The storage differences between SRO and deterministic case are not substantial due to the limited investment number of storage facilities, which can also be observed from the storage capital and operating costs in Fig. 5.

6.2. Sensitivity analysis based on SRO

In the following, we investigate the impacts of the number of clusters c and different uncertainty budget values Φ_{gtc} on the solution results. More clusters means more accurate approximation of the original demand data, but at the same time, it leads to more variables and constraints, which may be computationally intractable, especially when solving the MILP problem directly by Gurobi. As we can see from Table 1, the total running time is around 10 mins when 4 clusters are employed. However, when the cluster number increases to 12, the total running time increases exponentially. In Table 2, we show the model size under 12 clusters with $\Phi_{gtc} = 3$, and compare the running times and computational results by using monolithic way and our adopted hierarchical method shown in Section 4. The number of equations and continuous variables in the MILP problem is nearly tripled from 4 clusters to 12 clusters. Note that the number of variables fixed in the hierarchical solution process is also counted into the total number of variables when using GAMS. That is why the total number of variables in Step 1 and Step 2 remains the same. It can be known from Table 2 that the two-step hierarchical method significantly reduces the total running time with the number of clusters grows, while resulting in a slightly better solution than the monolithic way. Achieving a lower cost value with optimality gap smaller than 6.37% would require more running time for monolithic method due to its high combinatorial complexity.

Table 2
The comparison between two methods for SRO under 12 clusters and $\Phi_{gtc} = 3$.

	Monolithic	Hierarchical method	
		Step 1	Step 2
Disc. Var.	544	544	544
Cont. Var.	1,416,672	1,416,672	1,416,672
Equations	1,257,656	1,164,344	1,257,656
Optimality Gap (%)	6.37	2.06	3.12
CPU time	13 h 07 min	30 min	
Total cost (£b)	110.48	107.42	

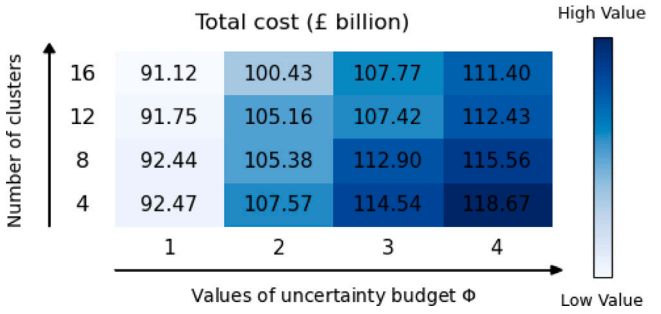


Fig. 11. Heat map of solution profile for SRO under different number of clusters ($c = 4, 8, 12, 16$) and uncertainty budget values ($\Phi_{gtc} = 1, 2, 3, 4$).

The uncertainty budget can affect the size of uncertainty set, i.e., the uncertainty data coverage, and thus affect the amount of hydrogen demand to be satisfied under the worst-case scenario. A larger value of uncertainty budget means more data coverage. In Fig. 11, we present a heat map for the SRO under different values of the parameters. It can be observed that increasing the value of uncertainty budget leads to a higher system cost, while increasing the number of clusters results in a lower cost because the weight of each uncertain representative day introduced in Section 3.1 will decrease as the number of clusters grows, and thus the total demand will decrease. Note that the total cost of 4 kinds of cluster numbers shown in Fig. 11 is very close when budget $\Phi_{gtc} = 1$ representing less data coverage because the worst-case demand of each cluster is close to its average value and then the obtained total demand after timing the weight of each cluster is similar as well. If we set Φ_{gtc} equal to its maximum value, i.e., the uncertain data dimension 24, the uncertainty set would become a box set, which will lead to a very conservative solution and could not reflect the real situation. In fact, when the value of Φ_{gtc} is 3 or 4, it already contains most of the uncertain data. Our results can provide a certain valuable reference on choosing the value of Φ_{gtc} for decision makers in real applications. Fig. 12 demonstrates a near linear relationship between running time and the expansion of cluster number, maintaining similar trend under different uncertainty budgets. When the number of clusters increases to 16, it only takes about 80 mins to get the solution of the large-scale MILP problem by using the hierarchical method.

6.3. Comparative analysis of SRO and ARO

The SRO does not provide feedback for uncertainty and all decisions are made "here and now". It needs to guarantee that the solution is feasible satisfying the energy balance constraint for every hour, which means that the worst-case scenario in the uncertainty set is taken at each moment. In practice, it is unlikely that all uncertain parameters will get the worst-case values at the same time. Compared with SRO, the decisions of ARO are made on the "wait and see" basis and part of decisions like the operational decisions can be adjusted after the uncertainty realisations. Therefore, the ARO can avoid the overly-conservative issue of SRO. In this subsection, we will analyse

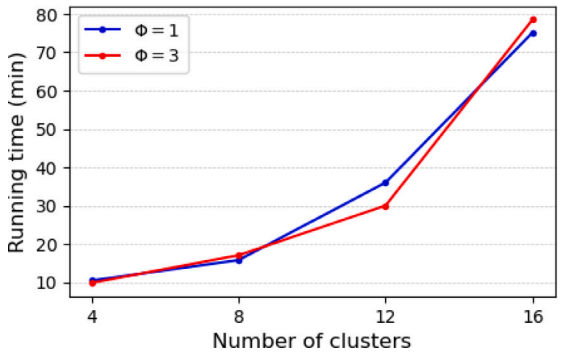


Fig. 12. Total running time for SRO under different number of clusters with $\Phi = 1$ and 3.

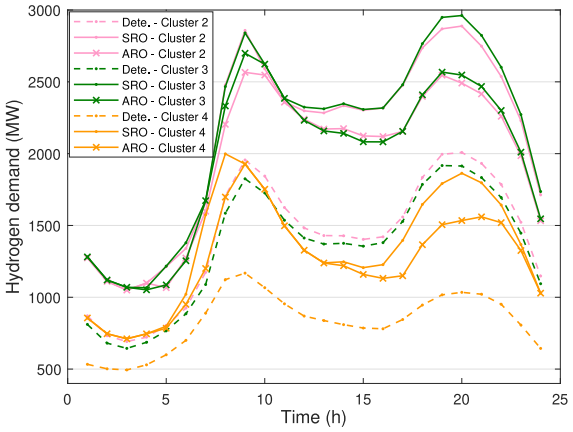


Fig. 13. The hydrogen demand comparison among deterministic (Dete.) and SRO and ARO under worst-case scenario with $\Phi = 2$ for each cluster of region SC in 2035 under 4 clusters.

the difference between ARO and SRO, and compare their performance on the hydrogen infrastructure planning problem under 4 clusters. We set the value of uncertainty budget $\Phi = 2$ in this part for showing different perspectives and comparison and avoiding presenting same repetitive results as in Section 6.1 for SRO. The ARO problem is solved by applying the developed two-level hybrid ECCG algorithm shown in Section 5. Fig. 13 displays the hydrogen demand of region SC under deterministic case, and the demand under worst-case scenario between SRO and ARO in 2035 under 4 clusters and $\Phi = 2$. Since the peak demand day is treated as one cluster, it only displays the daily profiles of Cluster 2, 3 and 4. It can be observed that the demand will increase under uncertainty compared with the deterministic case. The demand for SRO is higher than that for ARO at most hours within each cluster, especially after 13:00 p.m., because it adopts the maximum value of demand within the uncertainty set for each hour, as shown in Eq. (8) and the decisions of SRO are made at once before the uncertainty realisation, while ARO considers the system worst-case scenario for the whole 24 h rather than only for individual hour, as shown in Eq. (16). It is worth noting that ARO can incur lower total cost not only due to the lower demand under the worst-case scenario compared to SRO, but also because the system can adjust operational decisions based on the realised uncertainty.

In Table 3, we compare the total system cost and running time for SRO and ARO under 4 clusters and $\Phi = 2$. It can be seen that ARO results in a lower cost compared to SRO, and at the same time it needs more CPU execution time because SRO represents single stage implementation, while ARO involves two-stage operation with iterations.

Table 3
Cost comparison between ARO and SRO models under 4 clusters and $\Phi_{gtc} = 2$.

Models	Total cost (£b)	Running time
SRO	107.57	9.8 min
ARO	94.99	11 h 32 min

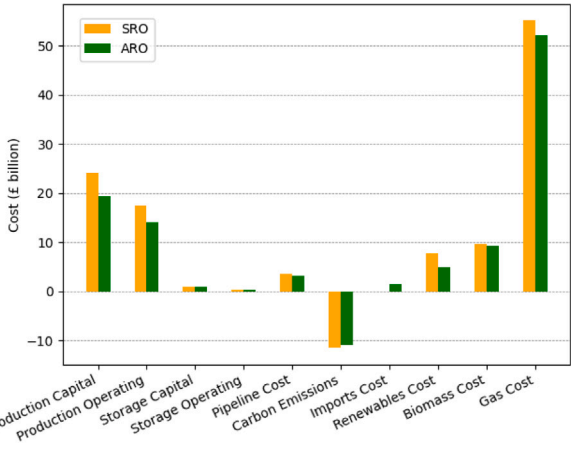


Fig. 14. Cost comparison between SRO and ARO models under 4 clusters and $\Phi = 2$.

Fig. 14 shows the detailed cost comparison between these two models. Most of the cost like production capital and operational cost and gas cost in ARO framework decreases in line with the reduced demand, while the import cost related to the operational decision import rate increases.

The convergence of proposed ECCG algorithm under 4 clusters and $\Phi = 2$ is displayed in Fig. 15. The X-axis and Y-axis represent the number of iterations and the total cost (objective value), respectively. As the algorithm iterates, the size of each CCG-SP remains the same while the size of the CCG-MP significantly increases because the optimality cuts are constantly added to the master problem at each iteration shown in (15). More specifically, CCG-MP has 544 discrete variables, 108,375 continuous variables, and 180,207 constraints at the first iteration, while it increases to 544 discrete variables, 215,901 continuous variables, and 358,046 constraints at the second iteration, and 544 discrete variables, 323,427 continuous variables, 535,885 constraints at the third iteration. Therefore, the computational time required for optimising the problem also increases progressively. As we can see, the proposed ECCG algorithm takes only 3 iterations to reach the optimality tolerance of 0.1%. The needed CPU times at the first and second iterations are within 14 mins, while it goes to around 11 h at the third iteration involving 5.47 h for Step 1 and 5.58 h for Step 2 in the large-scale CCG-MP solving process. If we solve the CCG-MP directly by using a monolithic approach rather than the hierarchical method, it would require significant computing time as the iteration step increases, especially under more clusters shown in Table 2. Furthermore, compared with the commonly used decomposition algorithm in the literature [53] where the CCG-MP is solved by the monolithic way and the CCG-SP is solved by using a big-M method with introduction of additional binary variables that amount to 1872 in our problem across all technologies, regions and years, the adopted BCD method in the ECCG algorithm only needs to solve two linear problems for CCG-SP and can converge within 20 s to the optimality criteria 1×10^{-8}, further improving the computational efficiency. In Fig. 16, we show the error value of the BCD method at outer iteration step $K = 3$, which only requires 5 iterations to reach the convergence criterion. In addition, the BCD method does not involve formulating the complex dual problem of the CCG-SP and suffering from the issue of appropriately selecting the bounds of the dual variables which are required by the big-M method.

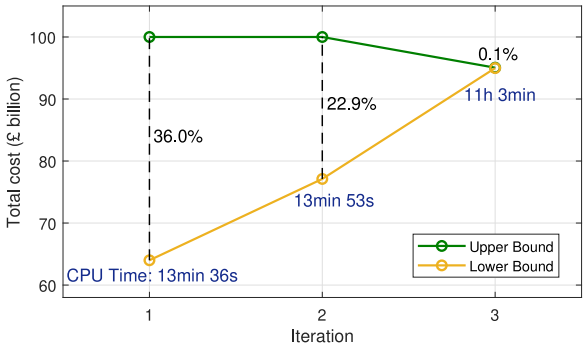


Fig. 15. The upper and lower bounds of outer ECCG algorithm for ARO under 4 clusters and $\Phi = 2$.

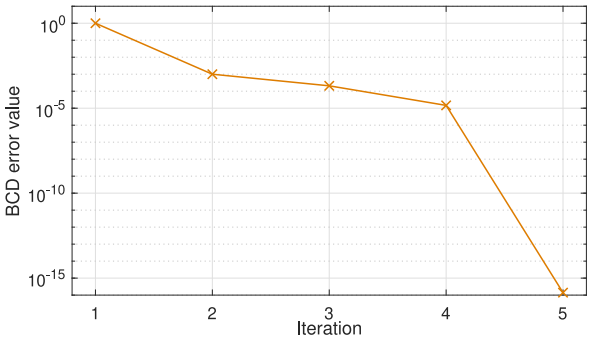


Fig. 16. The error value between upper and lower bounds of BCD algorithm (i.e., value of $|C^{IL} - \Phi_{lo}{}^{v}|/|C^{IL}|$) for ARO under 4 clusters and $\Phi = 2$ at outer iteration step $K = 3$.

In Fig. 17, we compare the total production capacity between SRO and ARO models among different technologies and years. The production of ATRCCS technology for ARO decreases compared to SRO in each year due to the decreasing demand, while the production of SMRCCS technology increases. Nevertheless, the ATRCCS is still the most prevailing technology. Compared with the deterministic case in Fig. 6, ARO and SRO both involve the WE production from the renewables. Fig. 18 shows the temporal and spatial heatmap of production capacity under these two frameworks. Compared with SRO, the production for ARO are mainly concentrated in several high-demand regions, especially NO, NE, WS and SE regions in 2045 and 2050. Differently from SRO where the production percentage across all regions except NO region is similar in each year, the production percentage for ARO varies significantly across most regions, which can also be observed from Fig. 19 showing the detailed percentage of production capacity for each region in 2050 year.

The temporal and spatial heatmap of storage capacity for SRO and ARO is displayed in Fig. 20. We can observe that the storage capacity for SRO in different regions of each year is very close, while the storage for ARO is concentrated in SC, NW, WM, EM, EA and NT regions and the amount of storage in these regions is much larger than SRO's. In Fig. 21, we compare the total storage between SRO and ARO in different years, and find that their difference is not substantial in each year. Based on these observations, it can be inferred that ARO tends to concentrate hydrogen storage in specific regions compared to SRO, subsequently transporting it to other areas through pipelines. The availability of hydrogen pipelines under the ARO framework from 2040 to 2050 (which remains constant during these years) is illustrated in Fig. 22 marking the pipeline transmission distance across regions (unit: km), which is similar to the pipeline availability in SRO. It can be observed that these storage-concentrated regions SC, NW, WM, EM, EA and NT are central regions, and hence the distances they

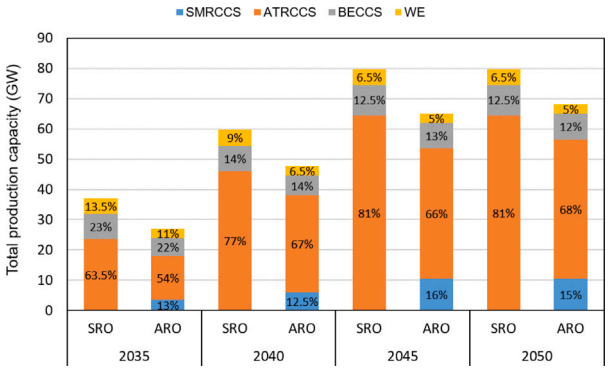


Fig. 17. The total production capacity comparison of different technologies between SRO and ARO models in different years under 4 clusters and $\Phi_{gtc} = 2$.

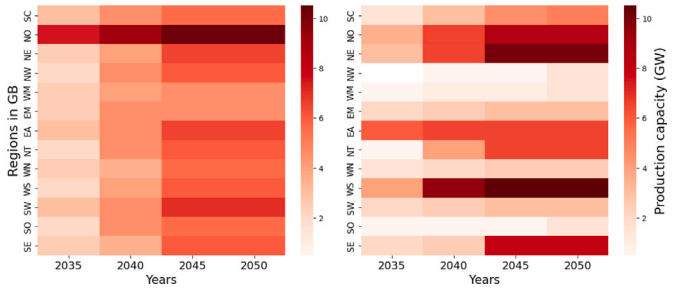
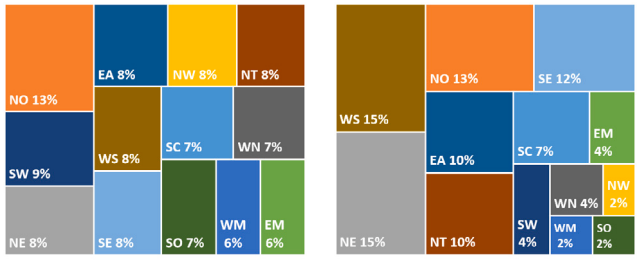


Fig. 18. Temporal and spatial heatmap of production capacity for SRO (left) and ARO (right) under 4 clusters and $\Phi_{gtc} = 2$.



(a) SRO (b) ARO

Fig. 19. The percentage of production capacity for each region in 2050 year for SRO and ARO under 4 clusters and $\Phi_{gtc} = 2$.

transport hydrogen to other regions are shorter. Additionally, other regions with surplus production, such as WS and SE, can also efficiently transmit hydrogen to these central areas. Furthermore, compared with the distributed storage, the centralised storage can lead to a lower storage investment because the surplus storage capacity in each region is effectively utilised and concentrated, and similarly it applies to production capacity as well. Since part of decisions of ARO can be adjusted to better accommodate the uncertainty realisations, the transmission operation of pipelines become more frequent than SRO and can reduce the system economic cost to some extent.

In an effort to further elucidate the degree of conservativeness of different solutions, i.e. Deterministic vs ARO and SRO, we conduct a set of optimisation runs during which we fix the investment decisions for generation, storage and transmission and allow only the operational decisions to be optimised. That way we aim to mimic the real-life implementation of the strategic decisions within the context of capacity expansion planning towards deep decarbonisation. In Fig. 23, we present a breakdown of the capital, operational and total costs. Note that the capital costs are fixed from the solution of integrated

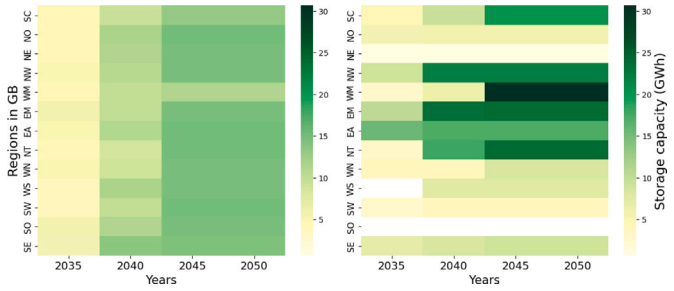


Fig. 20. Temporal and spatial heatmap of storage capacity for SRO (left) and ARO (right) under 4 clusters and $\Phi_{gtc} = 2$.

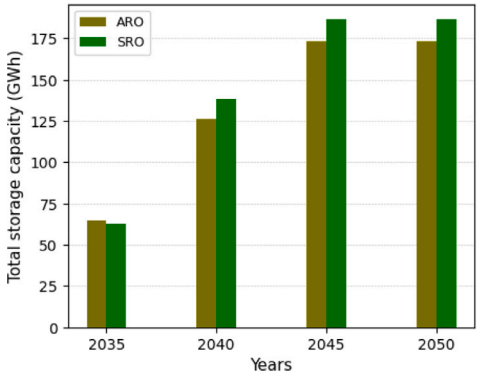


Fig. 21. Total storage capacity for ARO and SRO under 4 clusters and $\Phi_{gtc} = 2$.

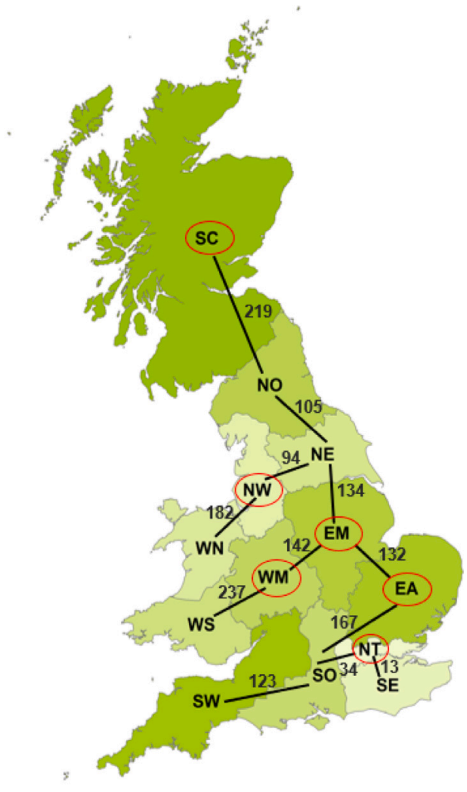


Fig. 22. Availability and distances (marked number with unit: km) across regions of hydrogen pipelines for ARO from 2040–2050 under 4 clusters and $\Phi_{gtc} = 2$.

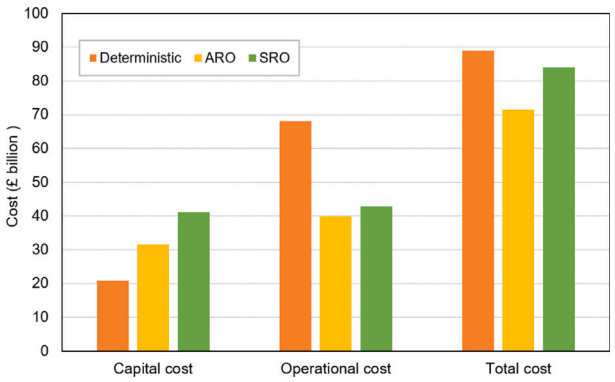


Fig. 23. Economic performance comparison between deterministic, ARO and SRO models when tested in a Monte Carlo simulation environment.

capacity and operational optimisation problem, while the operational cost is computed via the whole year operational optimisation with fixed capacity investments. In the full-year optimisation of operations, we allow for demand shedding with a value of lost load (VoLL) at £20,000/MWh, similar to the recent study [68]. Focusing on Fig. 23, it is evident that the most expensive strategy in terms of capital expenditures is the SRO by 97% and 30% more funds needed compared to the deterministic and ARO, respectively. Interestingly, when focusing on the system's operational expenditures a trend reversal is observed, where ARO solution results in 7% less costs compared to the SRO while the deterministic solution results in 59% more cost due to the inability to account for the uncertainty and resorting to demand shedding. In terms of the final total cost of the decarbonised system we see that the most expensive strategy when deployed in a "real-life" scenario is the deterministic approach by 25% while the least expensive approach and the one that appears as the best compromise is the ARO with 15% less overall costs when compared to SRO. We further conducted a sensitivity analysis for different VoLL values ranging from 25,000 to 40,000. The results show that general trend remains the same and the gap between the uncertainty-aware solutions and the deterministic one is gradually exacerbating. Overall, through this study, we highlight the urgent need to take into account uncertainty in the development of energy systems planning models with representative days formulation to secure informed, resilient and cost-effective decarbonisation strategies.

In this part, we have investigated how the planning of future hydrogen infrastructure is affected by intra-day demand variability, and demonstrated the cost and resilience differences between SRO and ARO methodologies. The one-shot investment option of SRO, while simpler to implement, results in capital losses. Conversely, ARO, with more frequent operations, tends to yield lower capacity expansion and reduced capital costs. Our results indicate that the majority of hydrogen from 2035–2050 continues to be provided by natural gas through ATR-CCS technology. In order to achieve long-term heat decarbonisation targets in GB, it will be necessary to invest renewable generation for hydrogen production, expand national-scale pipeline infrastructure for both H_2 and CO_2, and install large volumes of energy storage facilities. There are notable differences in the production and storage installation locations between SRO and ARO, more specifically, SRO tends to favour similar distribution across various regions, while ARO prefers centralised installation in specific regions. These insights can be used by national policymakers to assess which decision-making methodology is more appropriate to their use case.

7. Conclusions

In this paper, we proposed a single-level SRO model and a two-stage ARO model for the spatially-explicit multi-period large-scale hydrogen infrastructure planning problems under demand uncertainty. The demand uncertainty is captured by the introduction of novel concept of data-driven robust representative days. Our developed hybrid algorithm can reduce the computing time by 30%–90% compared to the monolithic way as the number of clusters grows while keeping a reasonable optimality gap. Compared with deterministic cases, the robust frameworks are more resistant to risks caused by uncertainty, while the extent of conservatism of solution can be controlled by the uncertainty budgets of polyhedral uncertainty sets. ATRCCS production technology is the most commonly used compared to other technologies due to its high energy conversion efficiency and its capacity takes up 50%–85% of the total system production capacity. ARO can alleviate the conservative issues and reduce the total cost by around 11% compared to SRO since its operational decisions like the hydrogen transmission flowrates could be made after the uncertainty realisations and its hydrogen demand under worst-case scenario is lower. We can also observe from the simulation results that the production and storage of ARO are mainly concentrated in several regions, the amount of transmission between pipelines increases and the transmission operation is more frequent compared with the distributed production and storage of SRO to hedge against the uncertainty.

Ongoing research in our group focuses on extending the proposed approach of uncertainty-aware representative days to the optimal sector-coupling planning problems towards net-zero [69] as well as the implementation of novel stochastic programming methods [70] for REG uncertainties, and further developing novel frameworks with longer-term energy storage properties based on the representative weeks and efficient solution methods to reduce the uncertain model's conservatism, while considering the terminal residual value of infrastructure over a 40 years planning horizon.

CRediT authorship contribution statement

Xu Zhou: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Margarita E. Efthymiadou:** Writing – review & editing, Validation, Software, Data curation. **Lazaros G. Papageorgiou:** Writing – review & editing, Validation, Funding acquisition. **Vassilis M. Charitopoulos:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Financial support from the Engineering & Physical Sciences Research Council (EPSRC), UK under the projects EP/T022930/1, EP/W003317/1 and EP/V051008/1, is gratefully acknowledged.

Appendix. Deterministic hydrogen infrastructure planning model

The deterministic hydrogen infrastructure planning model that is employed in our work was firstly introduced in [24]. In the following sections the detailed model formulation is given.

A.1. Total cost

The capital costs (PCC) and (SCC) depend on the number of newly invested hydrogen production plants (IP_{pgt}) of production technology p and the number of newly invested hydrogen storage sites (IS_{sgt}) of storage type s in region g at year t.

$$PCC = \sum_{p\in\mathcal{P}}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}} dfc_t \cdot pcc_{pt} \cdot cap_p^{\mathrm{P}} \cdot IP_{pgt} \tag{A.1}$$

$$SCC = \sum_{s\in\mathcal{S}}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}} dfc_t \cdot scc_{st} \cdot cap_s^{\mathrm{S}} \cdot IS_{sgt} \tag{A.2}$$

where dfc_t is the discount factor for capital cost at t, the parameters pcc_{pt}/scc_{st} and $cap_p^{\mathrm{P}}/cap_s^{\mathrm{S}}$ denote the capital cost of hydrogen production/storage of production type p/storage type s and their production/storage capacity, respectively.

The pipeline capital cost ($PLCC$) includes the hydrogen pipeline cost between regions and to storage caverns, CO_2 onshore pipeline cost and CO_2 offshore pipeline cost to reservoirs. The pipeline operating cost ($PLOC$) is assumed to be a certain proportion ($\delta, \overline{\delta}, \overline{\overline{\delta}}$) of capital cost ($PLCC$).

$$\begin{aligned} PLCC = &\sum_{t\in\mathcal{T}}\sum_{g,g'\in\mathcal{N}_{gg'}^{\mathrm{pipe}}} dfc_t \cdot pc \cdot D_{gg'}^{\mathrm{pipe}} \cdot Y_{gg't} \\ &+ \sum_{t\in\mathcal{T}}\sum_{g,sa\in GS_{gs}} dfc_t \cdot pc \cdot D_{g,sa}^{\mathrm{st}} \cdot Y_{g,sa,t}^{\mathrm{S}} \\ &+ \sum_{t\in\mathcal{T}}\sum_{g,g'\in\mathcal{N}_{gg'}} dfc_t \cdot \overline{pc} \cdot D_{gg'}^{\mathrm{pipe}} \cdot \overline{Y}_{gg't} \\ &+ \sum_{t\in\mathcal{T}}\sum_{g,r\in GR_{gr}} dfc_t \cdot \overline{\overline{pc}} \cdot D_{gr}^{\mathrm{res}} \cdot \overline{\overline{Y}}_{grt} \end{aligned} \tag{A.3}$$

$$\begin{aligned} PLOC = &\sum_{t\in\mathcal{T}}\sum_{g,g'\in\mathcal{N}_{gg'}^{\mathrm{pipe}}} \delta \cdot dfo_t \cdot crf \cdot pc \cdot D_{gg'}^{\mathrm{pipe}} \cdot AY_{gg't} \\ &+ \sum_{t\in\mathcal{T}}\sum_{g,sa\in GS_{gs}} \delta \cdot dfo_t \cdot crf \cdot pc \cdot D_{g,sa}^{\mathrm{st}} \cdot AY_{g,sa,t}^{\mathrm{S}} \\ &+ \sum_{t\in\mathcal{T}}\sum_{g,g'\in\mathcal{N}_{gg'}} \overline{\delta} \cdot dfo_t \cdot crf \cdot \overline{pc} \cdot D_{gg'}^{\mathrm{pipe}} \cdot \overline{AY}_{gg't} \\ &+ \sum_{t\in\mathcal{T}}\sum_{g,r\in GR_{gr}} \overline{\overline{\delta}} \cdot dfo_t \cdot crf \cdot \overline{\overline{pc}} \cdot D_{gr}^{\mathrm{res}} \cdot \overline{\overline{AY}}_{grt} \end{aligned} \tag{A.4}$$

where binary variables $Y_{gg't}(Y_{g,sa,t}^{\mathrm{S}})/\overline{Y}_{gg't}/\overline{\overline{Y}}_{grt}$ and $AY_{gg't}$ $(AY_{g,sa,t}^{\mathrm{S}})/\overline{AY}_{gg't}/\overline{\overline{AY}}_{grt}$ denote the new H_2/onshore CO_2/offshore CO_2 pipeline establishment from region g to region g' (storage cavern sa)/reservoir r, and their pipelines availability, respectively. The parameters $D_{gg'}^{\mathrm{pipe}}$, $D_{g,sa}^{\mathrm{st}}$ and D_{gr}^{res} are the distances through pipeline transmission from region g to region g', from g to underground storage cavern sa, and from g to CO_2 reservoir r, respectively. $\mathcal{N}_{gg'}$ denotes a set of total connections between neighbouring regions. $\mathcal{N}_{gg'}^{\mathrm{pipe}}$ is a subset of $\mathcal{N}_{gg'}$, encompassing neighbouring regions that are connectable via hydrogen pipelines. GS_{gs} denotes the sets of region g in which storage facility s is located. GR_{gr} denotes the sets of region g and reservoir r connections. dfo_t is the discount factor for operational cost. The parameters $pc, \overline{pc}, \overline{\overline{pc}}$ are the pipeline costs, while crf is the capital recovery factor.

The total operational cost (POC) of hydrogen production depends on the number of available hydrogen plants (NP_{pgt}) and the production rate (Pr_{pgcht}) of each production type p. The total operational cost (SOC) of hydrogen storage depends on the number of the available storage site (NS_{sgt}) and the hydrogen flowrate (Q_{gstch}^{I}) via pipeline to storage type s in region g.

$$\begin{aligned} POC = &\sum_{p\in\mathcal{P}}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}} dfo_t \cdot poc_{pt}^{\mathrm{F}} \cdot cap_p^{\mathrm{P}} \cdot NP_{pgt} \\ &+ \sum_{p\in\mathcal{P}}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{c\in C}\sum_{h\in\mathcal{H}} dfo_t \cdot WF_c \cdot poc_{pt}^{\mathrm{V}} \cdot Pr_{pgcht} \end{aligned} \tag{A.5}$$

$$\begin{aligned} SOC = &\sum_{s\in\mathcal{S}}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}} dfo_t \cdot soc_{st}^{\mathrm{F}} \cdot cap_s^{\mathrm{S}} \cdot NS_{sgt} \\ &+ \sum_{g\in\mathcal{G}}\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}}\sum_{c\in C}\sum_{h\in\mathcal{H}} dfo_t \cdot WF_c \cdot soc_{st}^{\mathrm{V}} \cdot Q_{gstch}^{\mathrm{I}} \end{aligned} \tag{A.6}$$

where parameters $poc_{pt}^{\mathrm{F}}/soc_{st}^{\mathrm{F}}$ and $poc_{pt}^{\mathrm{V}}/soc_{st}^{\mathrm{V}}$ are fixed and variable operational cost of hydrogen production/storage of production type p/storage type s, respectively. WF_c is the weight of day c introduced in Section 3.1.

The carbon emissions cost (CEC) correlates with the hydrogen production rate and carbon emissions cost (ct_t):

$$CEC = \sum_{p\in\mathcal{P}}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{c\in C}\sum_{h\in\mathcal{H}} WF_c \cdot dfo_t \cdot ct_t \cdot y_{pt}^{\mathrm{e}} \cdot Pr_{pgtch} \tag{A.7}$$

where y_{pt}^{e} denotes the coefficient of CO_2 emissions for production technology p at t.

The import cost (IIC) depends on the import price (p^{imp}) and the import rate (I_{gtch}):

$$IIC = \sum_{g\in Gimp_g}\sum_{t\in\mathcal{T}}\sum_{c\in C}\sum_{h\in\mathcal{H}} WF_c \cdot dfo_t \cdot p^{\mathrm{imp}} \cdot I_{gtch} \tag{A.8}$$

where $Gimp_g$ denotes the set of regions in which international import are feasible.

The total renewable cost (ReC) is related to the new installed capacity (IR_{egt}) and available capacity (NR_{egt}) of renewable technology e in region g and year t.

$$ReC = \sum_{e\in\mathcal{E}}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}} (dfc_t \cdot rc_{et} \cdot IR_{egt} + dfo_t \cdot ro_{et} \cdot NR_{egt}) \tag{A.9}$$

where rc_{et} and ro_{et} are the capital and operating costs of renewable e at t, respectively.

The total fuel cost (FC) for the natural gas and biomass consumption is as follows:

$$FC = \sum_{p\in\mathcal{P}}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{c\in C}\sum_{h\in\mathcal{H}} dfo_t \cdot c_t^p \cdot WF_c \cdot \frac{Pr_{pgtch}}{\eta_{pt}} \tag{A.10}$$

where c_t^p is the fuel (gas/biomass) price at t, and η_{pt} is the production efficiency of technology p.

A.2. Production constraints

The total number of available production plants (NP_{pgt}) for each production technology p and year t is related to its new invested plants number (IP_{pgt}), which is expressed below:

$$NP_{pgt} = NP_{pg,t-1} + IP_{pgt} - IP_{pg,t-(\frac{LTP_p}{n})} \qquad \forall p,g,t \tag{A.11}$$

where LTP_p is the lifetime of technology p.

The hydrogen production rate (Pr_{pgtch}), for each p,g,t,c,h, is limited by the capacity of infrastructures:

$$0 \le Pr_{pgtch} \le cap_p^{\max} \cdot NP_{pgt} \quad \forall p,g,t,c,h. \tag{A.12}$$

The operation of production plants also needs to satisfy the hourly ramp-up and ramp-down constraints (A.13) and (A.14):

$$Pr_{pgtch} - Pr_{pgtc,h-1} \le RU_p \cdot cap_p^{\mathrm{P}} \cdot NP_{pgt} \quad \forall p,g,t,c,h \tag{A.13}$$

$$Pr_{pgtc,h-1} - Pr_{pgtch} \le RD_p \cdot cap_p^{\mathrm{P}} \cdot NP_{pgt} \quad \forall p,g,t,c,h \tag{A.14}$$

where RU_p and RD_p are the ramp-up and ramp-down rates for technology p, respectively.

A.3. Hydrogen storage constraints

The total number of available storage units (NS_{sgt}) is defined by Eq. (A.15):

$$NS_{sgt} = NS_{sg,t-1} + IS_{sgt} - IS_{sg,t-(\frac{LTS_s}{a})}, \quad \forall\{s,g\} \in GS_{gs}, t \tag{A.15}$$

where IS_{sgt} denotes the newly invested hydrogen storage sites, and LTS_s is the lifetime of storage technology s.

The hydrogen storage state (St_{sgtch}) for each storage type s and g,t,c,h, is limited by an upper and lower bound, as expressed by Eq. (A.16):

$$cap_s^{\min} \cdot NS_{sgt} \le St_{sgtch} \le cap_s^{\max} \cdot NS_{sgt}, \quad \forall\{s,g\} \in GS_{gs}, t, c, h. \tag{A.16}$$

The hydrogen storage state (St_{sgtch}) is shown in Eq. (A.17), where $Q^{\mathrm{I}}_{gstch}/Q^{\mathrm{R}}_{sgtch}$ denote the hydrogen flowrate via pipeline from/to region g to/from storage s, for each s,g,t,c,h. The storage state at final hour $h = 24$ of the daily horizon is assumed to be equal to the initial storage state ($St^{\mathrm{init}}_{sgtc,1}$) with $h = 1$ to ensure the continuity of storage levels across consecutive days. This assumption captures the typical daily operational patterns, which fits well within the proposed model framework and reduces the computing complexity. The 24-hour cycle for energy storage has been widely adopted in the literature [71,72], especially in the large-scale planning problem with the representative days formulation [51,52].

$$\begin{aligned} St_{sgtch} &= St_{sgtc,h-1} + Q^{\mathrm{I}}_{gstch} - Q^{\mathrm{R}}_{sgtch} \quad \forall\{s,g\} \in GS_{gs}, t, c, h \\ St_{sgtc,24} &= St^{\mathrm{init}}_{sgtc,1} \qquad \forall\{s,g\} \in GS_{gs}, t, c. \end{aligned} \tag{A.17}$$

In addition, the storage and withdrawal rates are limited by the following constraints:

$$Q^{\mathrm{I}}_{gstch} \le Q_s^{\mathrm{Imax}} \cdot NS_{sgt} \qquad \forall\{s,g\} \in GS_{gs}, t, c, h \tag{A.18}$$

$$Q^{\mathrm{R}}_{gstch} \le Q_s^{\mathrm{Rmax}} \cdot NS_{sgt} \qquad \forall\{s,g\} \in GS_{gs}, t, c, h \tag{A.19}$$

where Q_s^{Imax} and Q_s^{Rmax} are the maximum storage and withdrawal rates for storage type s, respectively.

A.4. H_2 & CO_2 pipelines transmission

The available number of pipelines for H_2 transmission between regions and storage to caverns (sa), and available number of pipelines for onshore and offshore CO_2 transmission are defined as follows:

$$\begin{aligned} AY_{gg't} &= AY_{gg',t-1} + Y_{gg't} - Y_{gg',t-(\frac{LT^{\mathrm{pipe}}}{n})} \quad \forall\{g,g'\} \in \mathcal{N}^{\mathrm{pipe}}_{gg'}, t \\ AY^{\mathrm{S}}_{g,sa,t} &= AY^{\mathrm{S}}_{g,sa,t-1} + Y^{\mathrm{S}}_{g,sa,t} - Y^{\mathrm{S}}_{g,sa,t-(\frac{LT^{\mathrm{pipe}}}{n})} \quad \forall\{g,sa\} \in GS_{g,sa}, t \\ \overline{AY}_{gg't} &= \overline{AY}_{gg't} + \overline{Y}_{gg't} - \overline{Y}_{gg',t-(\frac{LT^{\mathrm{pipe}}}{n})} \quad \forall\{g,g'\} \in N_{gg'}, t \\ \overline{\overline{AY}}_{grt} &= \overline{\overline{AY}}_{gr,t-1} + \overline{\overline{Y}}_{grt} - \overline{\overline{Y}}_{gr,t-(\frac{LT^{\mathrm{pipe}}}{n})} \quad \forall\{g,r\} \in GR_{gr}, t \end{aligned} \tag{A.20}$$

where LT^{pipe} is the lifetime of pipelines.

The Eqs. (A.21) and (A.22) describe the pipelines availability balance between two regions.

$$AY_{gg't} = AY_{g'gt} \qquad \forall g,g' \in \mathcal{N}^{\mathrm{pipe}}_{gg'}, t \tag{A.21}$$

$$\overline{AY}_{gg't} = \overline{AY}_{g'gt} \qquad \forall g,g' \in \mathcal{N}_{gg'}, t \tag{A.22}$$

The H_2 flowrate ($Q_{gg'tch}$) from g to g', onshore CO_2 flowrate ($\overline{Q}_{gg'tch}$) and offshore CO_2 flowrate ($\overline{\overline{Q}}_{grtch}$) in the pipelines should satisfy the following maximum flowrate constraints:

$$Q_{gg'tch} \le q^{\mathrm{Hmax}} \cdot AY_{gg't} \qquad \forall g,g' \in \mathcal{N}^{\mathrm{pipe}}_{gg'}, t, c, h \tag{A.23}$$

$$\overline{Q}_{gg'tch} \le q^{\mathrm{Cmax}} \cdot \overline{AY}_{gg't} \qquad \forall g,g' \in \mathcal{N}_{gg'}, t, c, h \tag{A.24}$$

$$\overline{\overline{Q}}_{grtch} \le q^{\mathrm{Cmax}} \cdot \overline{\overline{AY}}_{grt} \qquad \forall\{g,r\} \in GR_{gr}, t, c, h \tag{A.25}$$

where q^{Hmax} and q^{Cmax} denote the maximum flowrates of pipelines for H_2 and CO_2, respectively.

A.5. Energy balance constraints

The hydrogen balance constraint (A.26) ensures that, in each region g, year t, cluster c and hour h, the total hydrogen generated by production plants (Pr_{pgtch}) and transmitted from other regions ($Q_{g'gtch}$) and imports (I_{gtch}), and withdrawn from storage site s in this region (Q^{R}_{sgtch}) can satisfy the total hydrogen demand and hydrogen energy transmitted to other regions ($Q_{gg'tch}$) and injected to storage units (Q^{I}_{gstch}).

$$\begin{aligned} &\sum_{p\in\mathcal{P}} Pr_{pgtch} + \sum_{g'\in\mathcal{N}^{\mathrm{pipe}}_{g'g}} Q_{g'gtch} + \sum_{s\in GS_{gs}} Q^{\mathrm{R}}_{sgtch} + I_{gtch} \\ &\ge \sum_{g'\in\mathcal{N}^{\mathrm{pipe}}_{gg'}} Q_{gg'tch} + \sum_{s\in GS_{gs}} Q^{\mathrm{I}}_{gstch} + D_{gtch}, \quad \forall g,t,c,h \end{aligned} \tag{A.26}$$

where D_{gtch} is the uncertain hydrogen demand introduced in Section 3.2. The hydrogen balance can also be observed from Fig. 2 (right).

The CO_2 mass balance constraint is expressed by Eq. (A.27). The left-hand side involves the onshore CO_2 flowrates ($\overline{Q}_{g'gtch}$) from other regions g' to region g, and the captured CO_2 which is related to production rate (Pr_{pgtch}) with a CO_2 capture coefficient (y^{c}_{pt}). The right-hand side also considers the offshore CO_2 flowrates ($\overline{\overline{Q}}_{grtch}$) from region g to reservoir r.

$$\sum_{g'\in\mathcal{N}_{g'g}} \overline{Q}_{g'gtch} + \sum_{p\in\mathcal{P}} y^{c}_{pt} Pr_{pgtch} = \sum_{g'\in\mathcal{N}_{gg'}} \overline{Q}_{gg'tch} + \sum_{r\in GR_{gr}} \overline{\overline{Q}}_{grtch} \quad \forall g,t,c,h \tag{A.27}$$

A.6. Renewable generation constraints

Denote the power production of renewable technology $e \in \mathcal{E}$ for each g,t,c,h by $\widehat{Pr}_{egtch}$, and its power curtailment by CL_{gtch}. The hydrogen produced by WE technology is expressed as:

$$Pr_{pgtch} = \eta^{\mathrm{we}}_t \Big(\sum_{e\in\mathcal{E}} \widehat{Pr}_{egtch} - CL_{gtch}\Big) \quad \forall p \in \{WE\}, g, t, c, h. \tag{A.28}$$

The power curtailment has an upper limit, as shown in Eq. (A.29):

$$\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}} CL_{gtch} \le cl \cdot \sum_{e\in\mathcal{E}}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}} \widehat{Pr}_{egtch}, \quad \forall c,h \tag{A.29}$$

The produced electricity ($\widehat{Pr}_{egtch}$) of each renewable technology e depends on its capacity (NR_{egt}) and its availability (AV_{egch}) which is pre-given, as described by Eq. (A.30).

$$\widehat{Pr}_{egtch} = AV_{egch} \cdot NR_{egt} \qquad \forall e \in \mathcal{E}, g, t, c, h. \tag{A.30}$$

The available renewable capacity at year t is shown in Eq. (A.31) where IR_{egt} denotes the new invested capacity.

$$NR_{egt} = NR_{eg,t-1} + IR_{egt} \qquad \forall e \in \mathcal{E}, g, t. \tag{A.31}$$

The installation of renewable farms is limited, which leads to inequality (A.32):

$$NR_{egt} \le la_{eg} \qquad \forall e \in \mathcal{E}, g, t \tag{A.32}$$

where la_{eg} is the capacity upper bound of renewable technology e in region g.

A.7. H_2 imports & fuel consumption constraints

The hydrogen import rate cannot exceed a certain upper bound (I^{Up}_{gtc}) for each import region $g \in Gimp_g, t, c, h$:

$$I_{gtch} \le I^{\mathrm{Up}}_{gtc}, \qquad \forall g \in Gimp_g, t, c, h. \tag{A.33}$$

The amount of biomass consumed by BGCCS technology to produce hydrogen is computed by equality (A.34) with efficiency η_{pt}. The limited amount of available biomass leads to Eq. (A.35):

$$V_{gt}^{\text{bio}} = \sum_{p\in\{BGCCS\}} \sum_{c\in C} \sum_{h\in\mathcal{H}} WF_c \cdot \frac{Pr_{pgtch}}{\eta_{pt}} \qquad \forall g,t \tag{A.34}$$

$$V_{gt}^{\text{bio}} \le ba_{gt} \qquad \forall g,t \tag{A.35}$$

where ba_{gt} is a maximum threshold of consumption in region g and year t.

A.8. CO_2 storage & emission constraints

The CO_2 storage (RI_{rt}) of reservoir r at year t is expressed by Eq. (A.36), which is related to the CO_2 flowrates to the reservoir.

$$RI_{rt} = RI_{r,t-1} + \tau \sum_{g\in GR_{gr}} \sum_{c\in C} \sum_{h\in\mathcal{H}} WF_c \cdot \overline{\overline{Q}}_{grtch}, \quad \forall r\in\mathcal{R},t \tag{A.36}$$

where τ is the duration of time periods (years). The absorption level is limited by an upper bound as described by Eq. (A.37), where cap_r^{R} represents the reservoir capacity.

$$RI_{rt} \le \sum_{g\in GR_{gr}} cap_r^{\text{R}} \cdot \overline{\overline{AY}}_{grt} \qquad \forall r\in\mathcal{R},t \tag{A.37}$$

The total CO_2 emissions (E_t) at year t for hydrogen production Pr_{pgtch} is calculated by Eq. (A.38). It should satisfy the allowed emission limit et_t, which is expressed by Eq. (A.39).

$$E_t = \sum_{p\in\mathcal{P}} \sum_{g\in\mathcal{G}} \sum_{c\in C} \sum_{h\in\mathcal{H}} WF_c \cdot y_{pt}^{e} \cdot Pr_{pgtch}, \qquad \forall t \tag{A.38}$$

$$E_t \le et_t, \qquad \forall t \tag{A.39}$$

References

[1] Lowes R, Woodman B. Disruptive and uncertain: Policy makers' perceptions on UK heat decarbonisation. Energy policy 2020;142:111494.

[2] Guo Y, Peng L, Tian J, Mauzerall DL. Deploying green hydrogen to decarbonize China's coal chemical sector. Nature Commun 2023;14(1):8104.

[3] Mac Dowell N, Sunny N, Brandon N, Herzog H, Ku AY, Maas W, et al. The hydrogen economy: A pragmatic path forward. Joule 2021;5(10):2524–9.

[4] Net zero by 2050 - a roadmap for the global energy sector. 2021, URL www.iea.org/t&c/.

[5] De Mel I, Bierkens F, Liu X, Leach M, Chitnis M, Liu L, et al. A decision-support framework for residential heating decarbonisation policymaking. Energy 2023;268:126651.

[6] Charitopoulos VM, Fajardy M, Chyong CK, Reiner DM. The impact of 100% electrification of domestic heat in Great Britain. Iscience 2023;26(11):1–12.

[7] National Grid ESO, Future energy scenarios. 2022, URL www.nationalgrideso.com/future-energy/future-energy-scenarios.

[8] BEIS, UK hydrogen strategy. 2021, https://www.gov.uk/government/publications/uk-hydrogen-strategy.

[9] Kakodkar R, He G, Demirhan C, Arbabzadeh M, Baratsas S, Avraamidou S, et al. A review of analytical and optimization methodologies for transitions in multi-scale energy systems. Renew Sustain Energy Rev 2022;160:112277.

[10] Moreno-Benito M, Agnolucci P, Papageorgiou LG. Towards a sustainable hydrogen economy: Optimisation-based framework for hydrogen infrastructure development. Comput Chem Eng 2017;102:110–27.

[11] Aunedi M, Yliruka M, Dehghan S, Pantaleo AM, Shah N, Strbac G. Multi-model assessment of heat decarbonisation options in the UK using electricity and hydrogen. Renew Energy 2022;194:1261–76.

[12] Sunny N, Mac Dowell N, Shah N. What is needed to deliver carbon-neutral heat using hydrogen and CCS? Energy Environ Sci 2020;13:4204–24.

[13] Reuß M, Grube T, Robinius M, Stolten D. A hydrogen supply chain with spatial resolution: Comparative analysis of infrastructure technologies in Germany. Appl Energy 2019;247:438–53.

[14] Zhang J, Li Z, Zheng X, Liu P. Long-term planning and coupling optimization of multi-regional natural gas and hydrogen supply systems: A case study of China. Comput Chem Eng 2024;183:1–14.

[15] He G, Mallapragada DS, Bose A, Heuberger-Austin CF, Gençer E. Sector coupling via hydrogen to lower the cost of energy system decarbonization. Energy Environ Sci 2021;14(9):4635–46.

[16] Ogumerem GS, Kim C, Kesisoglou I, Diangelakis NA, Pistikopoulos EN. A multi-objective optimization for the design and operation of a hydrogen network for transportation fuel. Chem Eng Res Des 2018;131:279–92.

[17] Pérez-Uresti SI, Gallardo G, Varvarezos DK. Strategic investment planning for the hydrogen economy - A mixed integer non-linear framework for the development and capacity expansion of hydrogen supply chain networks. Comput Chem Eng 2023;179:1–15.

[18] Güler MG, Geçici E, Erdoğan A. Design of a future hydrogen supply chain: A multi period model for Turkey. Int J Hydrog Energy 2021;46(30):16279–98.

[19] Cantú VH, Ponsich A, Azzaro-Pantel C, Carrera E. Capturing spatial, time-wise and technological detail in hydrogen supply chains: A bi-level multi-objective optimization approach. Appl Energy 2023;344:121159.

[20] Almaraz SD-L, Rácz V, Azzaro-Pantel C, Szántó ZO. Multiobjective and social cost-benefit optimisation for a sustainable hydrogen supply chain: Application to Hungary. Appl Energy 2022;325:119882.

[21] Ehrenstein M, Galán-Martín Á, Tulus V, Guillén-Gosálbez G. Optimising fuel supply chains within planetary boundaries: A case study of hydrogen for road transport in the UK. Appl Energy 2020;276:115486.

[22] Pastore LM, Groppi D, Feijoo F, Basso GL, Garcia DA, de Santoli L. Optimal decarbonisation pathways for the Italian energy system: Modelling a long-term energy transition to achieve zero emission by 2050. Appl Energy 2024;367:123358.

[23] Bui M, Zhang D, Fajardy M, Mac Dowell N. Delivering carbon negative electricity, heat and hydrogen with BECCS–Comparing the options. Int J Hydrog Energy 2021;46(29):15298–321.

[24] Efthymiadou ME, Charitopoulos VM, Papageorgiou LG. Optimal hydrogen infrastructure planning for heat decarbonisation. Chem Eng Res Des 2024;204:121–36.

[25] Riera JA, Lima RM, Knio OM. A review of hydrogen production and supply chain modeling and optimization. Int J Hydrog Energy 2023;48:13731–55.

[26] Uusitalo L, Lehikoinen A, Helle I, Myrberg K. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environ Model Softw 2015;63:24–31.

[27] Feng W, Ruiz C. Risk management of energy communities with hydrogen production and storage technologies. Appl Energy 2023;348:121494.

[28] Birge JR, Louveaux F. Introduction to Stochastic Programming. Springer Science & Business Media; 2011.

[29] Shang C, Huang X, You F. Data-driven robust optimization based on kernel learning. Comput Chem Eng 2017;106:464–79.

[30] Jabr RA. Robust transmission network expansion planning with uncertain renewable generation and loads. IEEE Trans Power Syst 2013;4(28):4558–67.

[31] Minguez R, Garcia-Bertrand R, Arroyo JM, Alguacil N. On the solution of large-scale robust transmission network expansion planning under uncertain demand and generation capacity. IEEE Trans Power Syst 2018;33(2):1242–51.

[32] Roldan C, Minguez R, Garcia-Bertrand R, Arroyo JM. Robust transmission network expansion planning under correlated uncertainty. IEEE Trans Power Syst 2019;34(3):2071–82.

[33] Riepin I, Schmidt M, Baringo L, Müsgens F. Adaptive robust optimization for European strategic gas infrastructure planning. Appl Energy 2022;324:1–19.

[34] Cai P, Mi Y, Ma S, Li H, Li D, Wang P. Hierarchical game for integrated energy system and electricity-hydrogen hybrid charging station under distributionally robust optimization. Energy 2023;283:1–15.

[35] Zhang Y, Liu F, Wang Z, Su Y, Wang W, Feng S. Robust scheduling of virtual power plant under exogenous and endogenous uncertainties. IEEE Trans Power Syst 2021;37(2):1311–25.

[36] Nunes P, Oliveira F, Hamacher S, Almansoori A. Design of a hydrogen supply chain with uncertainty. Int J Hydrog Energy 2015;40(46):16408–18.

[37] Kim J, Lee Y, Moon I. Optimization of a hydrogen supply chain under demand uncertainty. Int J Hydrog Energy 2008;33(18):4715–29.

[38] Almansoori A, Shah N. Design and operation of a stochastic hydrogen supply chain network under demand uncertainty. Int J Hydrog Energy 2012;37(5):3965–77.

[39] Dayhim M, Jafari MA, Mazurek M. Planning sustainable hydrogen supply chain infrastructure with uncertain demand. Int J Hydrog Energy 2014;39(13):6789–801.

[40] Hwangbo S, Lee I-B, Han J. Mathematical model to optimize design of integrated utility supply network and future global hydrogen supply network under demand uncertainty. Appl Energy 2017;195:257–67.

[41] Ochoa Bique A, Maia LK, Grossmann IE, Zondervan E. Design of hydrogen supply chains under demand uncertainty–a case study of passenger transport in Germany. Phys. Sci. Rev. 2023;8(6):741–62.

[42] Kim S, Park J, Chung W, Adams D, Lee JH. Techno-economic analysis for design and management of international green hydrogen supply chain under uncertainty: An integrated temporal planning approach. Energy Convers Manage 2024;301:118010.

[43] Jang J, Lee H. Effective hydrogen supply chain management framework considering nonlinear multi-stage process uncertainties. Appl Energy 2024;367:123328.

[44] Du Y, Jia D, Li X. A robust dynamic hydrogenation network layout model considering station capacity expansion and flexible transportation modes. Int J Hydrog Energy 2024;58:223–38.

[45] Lv X, Li X, Xu C. A robust optimization model for capacity configuration of PV/battery/hydrogen system considering multiple uncertainties. Int J Hydrog Energy 2023;48(21):7533–48.

[46] Deng M, Bian B, Zhou Y, Ding J. Distributionally robust production and replenishment problem for hydrogen supply chains. Transp Res E 2023;179:1–23.
[47] Qiu Y, Li Q, Wang T, Yin L, Chen W, Liu H. Optimal planning of cross-regional hydrogen energy storage systems considering the uncertainty. Appl Energy 2022;326:1–12.
[48] Xu J, Li Q, Li J, Zhang ZH, Zheng L. Robust Design of a Green Hydrogen Supply Chain Under Spatiotemporal Imbalance. IEEE Trans Autom Sci Eng 2024;1–15.
[49] Eskandari M, Gilani H, Sahebi H, Sharifi S. Design and planning of global sustainable bio-hydrogen supply chain with uncertainty : A transportation-oriented robust model. Chem Eng Sci 2024;283:1–17.
[50] Frew BA, Becker S, Dvorak MJ, Andresen GB, Jacobson MZ. Flexibility mechanisms and pathways to a highly renewable US electricity future. Energy 2016;101:65–78.
[51] Lara CL, Mallapragada DS, Papageorgiou DJ, Venkatesh A, Grossmann IE. Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm. European J Oper Res 2018;271(3):1037–54.
[52] Lara CL, Siirola JD, Grossmann IE. Electric power infrastructure planning under uncertainty: stochastic dual dynamic integer programming (SDDiP) and parallelization scheme. Optim Eng 2020;21:1243–81.
[53] Ning C, You F. Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods. Comput Chem Eng 2018;112:190–210.
[54] Vaes J, Charitopoulos VM. A data-driven uncertainty modelling and reduction approach for energy optimisation problems. Comput Aided Chem Eng 2023;52:1161–7.
[55] Abedinia O, Zareinejad M, Doranehgard MH, Fathi G, Ghadimi N. Optimal offering and bidding strategies of renewable energy based large consumer using a novel hybrid robust-stochastic approach. J Cleaner Prod 2019;215:878–89.
[56] Simoes S, Zeyringer M, Mayr D, Huld T, Nijs W, Schmidt J. Impact of different levels of geographical disaggregation of wind and PV electricity generation in large energy system models: A case study for Austria. Renew Energy 2017;105:183–98.
[57] Zeyringer M, Price J, Fais B, Li PH, Sharp E. Designing low-carbon power systems for Great Britain in 2050 that are robust to the spatiotemporal and inter-annual variability of weather. Nature Energy 2018;3(5):395–403.
[58] Aunedi M, Pantaleo AM, Kuriyan K, Strbac G, Shah N. Modelling of national and local interactions between heat and electricity networks in low-carbon energy systems. Appl Energy 2020;276:1–18.
[59] Lappas NH, Gounaris CE. Multi-stage adjustable robust optimization for process scheduling under uncertainty. AIChE J 2016;62(5):1646–67.
[60] Gurobi Optimization, LLC, Houston, TX. 2023, URL https://www.gurobi.com.
[61] Aghamohamadi M, Mahmoudi A, Ward JK, Haque MH, Catalao JP. A block-coordinate-descent robust approach to incentive-based integrated demand response in managing multienergy hubs with must-run processes. IEEE Trans Ind Appl 2022;58(2):2352–68.
[62] Baringo L, Baringo A. A stochastic adaptive robust optimization approach for the generation and transmission expansion planning. IEEE Trans Power Syst 2017;33(1):792–802.
[63] Daneshvar M, Mohammadi-Ivatloo B, Zare K, Asadi S. Two-stage robust stochastic model scheduling for transactive energy based renewable microgrids. IEEE Trans Ind Inf 2020;16(11):6857–67.
[64] Zeng B, Zhao L. Solving two-stage robust optimization problems using a column-and- constraint generation method. Oper Res Lett 2013;41(5):457–61.
[65] Allen RC, Avraamidou S, Butenko S, Pistikopoulos EN. Solution strategies for integrated distribution, production, and relocation problems arising in modular manufacturing. European J Oper Res 2024;314(3):963–79.
[66] Tsimopoulos EG, Georgiadis MC. Strategic offers in day-ahead market co-optimizing energy and reserve under high penetration of wind power production: An MPEC approach. AIChE J 2019;65(7):1–11.
[67] BEIS, Hydrogen production costs. 2021, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1011506/Hydrogen_Production_Costs_2021.pdf.
[68] van Zuijlen B, Zappa W, Turkenburg W, van der Schrier G, van den Broek M. Cost-optimal reliable power generation in a deep decarbonisation future. Appl Energy 2019;253:113587.
[69] Bounitsis GL, Charitopoulos V. The value of ammonia towards integrated power and heat system decarbonisation. Sustain Energy Fuels 2024.
[70] Bounitsis GL, Papageorgiou LG, Charitopoulos VM. Stable optimisation-based scenario generation via game theoretic approach. Comput Chem Eng 2024;108646.
[71] Tostado-Véliz M, Jordehi AR, Fernández-Lobato L, Jurado F. Robust energy management in isolated microgrids with hydrogen storage and demand response. Appl Energy 2023;345:121319.
[72] Wu X, Zhao W, Li H, Liu B, Zhang Z, Wang X. Multi-stage stochastic programming based offering strategy for hydrogen fueling station in joint energy, reserve markets. Renew Energy 2021;180:605–15.