# Supplementary Information for

## What is needed to deliver carbon-neutral heat using hydrogen and CCS?

Nixon Sunny,<sup>ac</sup> Niall Mac Dowell<sup>ab</sup> and Nilay Shah<sup>ac</sup>

<sup>a</sup> Centre for Process Systems Engineering, Imperial College London, London
 <sup>b</sup> Centre for Environmental Policy, Imperial College London, London

<sup>c</sup> Department of Chemical Engineering, Imperial College London, London

## Contents

1	Mo	del components	<b>2</b>						
	1.1	Nomenclature	2						
	1.2	Equations	2						
	1.3	Discretisation of space	6						
	1.4	Discretisation of time	6						
	1.5	Input summary	7						
	1.6	Output summary	7						
	1.7	Modelling assumptions	8						
<b>2</b>	Methodological innovations								
	2.1	Hierarchical analysis: Time	8						
	2.2	Transition pathway	9						
	2.3	Statistical analysis	0						
3	Sup	plementary results 1	1						
	$3.1^{-}$	Representation of time	1						
	3.2	Peak day operating mix for $H_2$ production	2						
	3.3	Statistical analysis of the regression models	2						
	3.4	Impact of grid $O_2$ intensities	3						
	3.5	Impact of biomass availability and $CO_2$ intensity $\ldots \ldots \ldots$	5						

Sets and Indices	Description
$g,g'\in G$	Spatial grid cells/ nodes that are indexed (54)
$r \in R$	All resources (Natural gas, domestic heat, industrial heat, electricity,
	biomass, $H_2$ at 5, 20, 40, 60 and 80 bar, emitted $CO_2$ , dense-phase
	$CO_2)$
$ir \subseteq R$	Importable resources (Natural gas, electricity, biomass)
$sr \subseteq R$	All resources which can be stored as opposed to being emitted (Dense-
	phase $CO_2$ , $H_2$ at 60 and 80 bar)
$t, t' \in T$	Minor time periods $(15)$
$tm, tm' \in TM$	Major time periods (7)
$m \in M$	Performance metrics (CapEx, OpEx)
$j \in J$	Technologies (SMR with syngas $CO_2$ capture, SMR with flue gas cap-
	ture, ATR with CCS, ATR with GHR and CCS, Biomass gasification
	with CCS, Water electrolysis, smaller scale SMRs and ATRs, domes-
	tic and industrial natural gas, $H_2$ boiler, intraday cavern, interseasonal
	cavern, injection well, $H_2$ and $CO_2$ compressors)
$pj \subseteq J$	Process technologies
$sj \subseteq J$	Storage technologies (intraday cavern, interseasonal cavern, injection
	well)
$d \in D$	Distribution technologies (Pipelines - $18$ , $24$ , $36$ and $48$ inches for $H_2$ ,
	12 and 26 inches $CO_2$ pipes for both onshore and offshore transport)
$od \in D$	Onshore transportation/ distribution technologies
$dmo \in DMO$	Distribution modes $(e.g., flows at specific pressure ranges)$

Table 1: A complete list of the sets and indices used in the H<sub>2</sub>-CCS infrastructure design model.

## 1 Model components

### 1.1 Nomenclature

Tables 1, 2 and 3 summarise the indices, sets, parameters and variables used within this optimisation framework. Units of measurement are omitted here as many of the parameters are block parameters which are created out of many parameters with specific units.

### 1.2 Equations

The equations presented here should be combined with those presented in the paper to produce a complete list, which can be used as a general framework. The import rates are bounded by a maximum import rate placed on importable resources in importable locations as shown in equation (1). This is written as follows:

$$IM_{ir,q,t,tm} \le IM_{ir,q}^{MAX} \qquad \forall \ ir,g,t,tm \tag{1}$$

where  $IM_{ir,g}^{MAX}$  denotes the maximum import rates of an importable resource *ir* in a grid cell, *g*. Similarly, the flowrate of a resource is bounded by the capacity of their units as shown in equation (2):

$$\sum_{dmo} Q_{g,g',d,dmo,t,tm} \le N_{g,g',d,tm}^D \ Q_d^{MAX} \qquad \forall \ g,g',d,t,tm$$
(2)

where  $N_{g,g',d,t}^D$  is the number of distribution units used to transport a resource from grid cell g to g' using distribution technology d at major time period tm,  $Q_d^{MAX}$  is the maximum flowrate achievable using a single distribution unit for transporting a resource using distribution technology d.

Parameter	Description
CI	Carbon intensity of the heat supply
$DEM^{TOT}$	Total annual heating demand
$CRF_i$	Captial recovery factor for the different technologies
$ON_{a}$	Binary parameter to indicate if a grid, $g$ is onshore/ offshore
$INIT_{j,g}$	Initial capacity of technology, $j$ in grid cell, $g$ at the start of the planning horizon
$D_{r,g,t,tm}$	Demand for resource, $r$ in grid cell, $g$ at minor time, $t$ and at major time, $tm$
$NPC_j$	Nameplate capacity for technology type, $j$ (e.g., unit capacity of the technologies)
$ u_{g,g'}$	Distance between grid cells, $g$ and $g'$ , where $g \neq g'$
$OT_t$	Operational time corresponding to each time period, $t$
$Q_d^{MAX}$	Maximum flow rate of a resource through a single unit of distribution technology type, d
$IM_{ir,a}^{MAX}$	Maximum import rate of an importable resource, $ir$ in a grid, $g$
$\beta_{r,dmo,d}$	Conversion coefficient of a resource, $r$ in a distribution technology, $d$
	through a distribution mode, dmo
$UE_r$	Upstream emissions pertaining to a unit of the resource, $r$
$\mu_{pj,r}$	Production/ consumption rate coefficients for resource, $r$ using a process
	technology type, $pj$
$ au_{j,g,m}$	Coefficients related to the effects of an investment in a technology type, $\boldsymbol{j}$
	in grid cell, $g$ on the performance metric type, $m$
$NC_{d,m}$	Coefficients related to the effects of installation of a network using a dis-
	tribution technology type, $d$ on the performance metric type, $m$ (e.g.,
	CAPEX for construction of a $H_2$ pipeline)
$PC_{pj,m}$	Coefficients related to the effects of the processing of resources using a
	process technology type, $pj$ on the performance metric type, $m$
$QC_{d,m}$	Coefficients related to the effects of the flow of a resource using distribu-
	tion technology type, $d$ on performance metric type, $m$
$IMC_{ir,m}$	Coefficients related to the effects of importing a resource, $ir$ on performance metric type, $m$
$SC_{si.m}$	Coefficients related to the effects of using a storage technology, $sj$ to
- ; ; · · ·	store a resource on the performance metric type, $m$
$S^{MAX}_{sia}$	Maximum storage capacity of a storage technology, $sj$ in grid cell, $q$
$\Gamma^{SMAX}_{srsi}$	Maximum injection rate of a storable resource type, sr into a storage
01,01	technology type, sj
$\Gamma^{RMAX}_{an ai}$	Maximum retrieval rate of a storable resource type, $sr$ from a storage
51,83	technology type, sj
$OW_{m,tm}$	Overall objective weighting factor corresponding to performance metric
• 110,0110	type, $m$ in major time period, $tm$

Table 2: A complete list of the parameters used in the  $H_2$ -CCS infrastructure design model.

Table 3: A	. complete l	ist of the	variables	used in	the $H_2$ -CCS	infrastructure	design mode	1.
------------	--------------	------------	-----------	---------	----------------	----------------	-------------	----

Variable	Description
$P_{pj,g,t,tm}$	Production rate of a process technology type, $pj$ operating in grid
	cell, $g$ at minor time, $t$ and major time, $tm$
$IM_{r,g,t,tm}$	Import rate of a resource type, $r$ in grid cell, $g$ at minor time, $t$ and
	major time, $tm$
$Q_{q,q',d,dmo,t,tm}$	Flowrate of resource from grid cell, $g$ to $g'$ , through distribution
	mode, $dmo$ in a distribution technology, $d$ at minor time, $t$ and ma-
	jor time, $tm$
$\varepsilon_{r,g,t,tm}$	Emission rate of resource type, $r$ in grid cell, $g$ at minor time, $t$ and
	major time, $tm$ (Intermediates that aren't stored are emitted.)
$\lambda_{r.a.sj.t.tm}^S$	Storage rate of resource type, $r$ using storage technology, $sj$ in grid
	cell, $g$ at minor time, $t$ and major time, $tm$
$\lambda_{r,q,sj,t,tm}^R$	Retrieval rate of resource type, $r$ from storage technology, $sj$ in grid
	cell, $g$ at minor time, $t$ and major time, $tm$
$I_{r,g,sj,t,tm}$	Total inventory of a resource type, $r$ using storage technology type, $sj$
	in grid cell, $g$ at minor time, $t$ and major time, $tm$
$TM_{m,tm}$	Total contribution to the performance metric type, $m$ in each major
	time period, $tm$
TAC	Total annualised costs of the entire network
$N^P_{pj,g,tm}$	Number of process units of process technology type, $pj$ existing in
~	grid cell, $g$ at major time period, $tm$
$N_{r,sj,g,tm}^S$	Number of storage units of storage technology type, $sj$ used to store
5	resource type, $r$ in grid cell, $g$ at major time period, $tm$
$N^{D}_{g,g',d,tm}$	Number of units of distribution technology type, $d$ operational from
D	grid cell, $g$ to $g'$ at major time period, $tm$
$NI_{pj,g,tm}^P$	Number of additional process units of process technology type, $pj$
C	built in grid cell, $g$ at major time period, $tm$
$NI^{\mathcal{S}}_{r,sj,g,tm}$	Number of additional storage units of storage technology type, $sj$
	used to store resource type, $r$ built in grid cell, $g$ at major time pe-
D	riod, <i>tm</i>
$NI^{D}_{g,g',d,tm}$	Number of additional units of distribution technology type, $d$ built
	from grid cell, $g$ to $g'$ at major time period, $tm$
$lpha_{g,tm}$	Binary indicating if the heat demand in a grid cell, $g$ is supplied using
	$H_2$ at major time period, $tm$

Further constraints are used to ensure that network is not built between a grid cell q to the exact same grid cell in equation (3), where g' = g. This in combination with equation (2) ensures that a flow cannot occur between the same grid cells. In addition, equation (4) ensures that at most one pipeline of a particular type is allowed to be built between any two grid cells. Equation (3) and (4) are written as follows:

$$N_{g,g',d,tm}^D = 0 \qquad \forall \ g,d,t,g' = g \tag{3}$$

$$N^{D}_{g,g',d,tm} \le 1 \qquad \forall \ g,g',tm,d \tag{4}$$

Equation (5) imposes constraints on the inventory of a resource as a result of the number of storage units and technologies available as shown below:

$$I_{r,g,sj,t,tm} \le N_{r,sj,g,tm}^S \ NPC_{sj} \qquad \forall \ r,g,sj,t,tm$$
(5)

$$I_{r,g,sj,t,tm} = I_{r,g,sj,t-1,tm} + OT_t(\lambda_{r,g,sj,t,tm}^S - \lambda_{r,g,sj,t,tm}^R)$$

$$\forall r, q, sj, t, tm$$
(6)

where  $I_{r,g,sj,t,tm}$  is the inventory of resource r using storage technology sj in grid cell g at minor time period t, and major time period, tm.  $N_{r,sj,g,tm}^S$  is the number of storage units storing resource r using technology sj in grid cell g at major time tm.  $OT_t$  is the operating time corresponding to the minor time period, t. It is assumed that the inventory is updated at the end of each time period. The equation is modified for when t reaches the last ordered element of the set T, where the inventory value from within that major time, tm is used to update the initial inventory for the next tm.

Equation (7) ensures that the inventory of a resource never exceeds the maximum theoretical capacity of the sites in which it is stored. This equation applies to inventories in all underground locations.

$$I_{r,g,sj,t,tm} \le S_{sj,q}^{MAX} \qquad \forall r,g,sj,t,tm$$

$$\tag{7}$$

where  $S_{sj,g}^{MAX}$  is the maximum storage capacity available via storage technology, sj in grid cell, g. The time-period linking unit balance constraints for storage and distribution units are shown in equations (8) and (9) respectively:

$$N_{r,si\,a\,tm}^{S} = N_{r,si\,a\,tm-1}^{S} + NI_{r,si\,a\,tm}^{S} \qquad \forall r,sj,g,tm$$

$$\tag{8}$$

$$N_{r,sj,g,tm}^{S} = N_{r,sj,g,tm-1}^{S} + NI_{r,sj,g,tm}^{S} \qquad \forall \ r, sj, g, tm$$

$$N_{g,g',d,tm}^{D} = N_{g,g',d,tm-1}^{D} + NI_{g,g',d,tm}^{D} \qquad \forall \ g, g', d, tm$$
(8)
(9)

where  $NI_{r,sj,g,tm}^S$  and  $NI_{g,g',d,tm}^D$  are the additional number of storage and distribution units built in major time period tm respectively. Constraints on storage and retrieval rates enforce that you cannot charge into and discharge from the same stores at the same time as shown in equations (10) and (11):

$$\lambda_{r,g,sj,t,tm}^S \le \Gamma_{r,sj}^{SMAX} \qquad \forall \ r,g,sj,t,tm$$
(10)

$$\lambda_{r,g,sj,t,tm}^R \le \Gamma_{r,sj}^{RMAX} \qquad \forall \ r,g,sj,t,tm \tag{11}$$

where  $\Gamma_{r,sj}^{SMAX}$  is the maximum injection rate of a resource r into a storage technology sj and  $\Gamma_{r,sj}^{RMAX}$  is the maximum retrieval rate. Additional constraints are enforced to ensure that CO<sub>2</sub> is never retrieved from the storage sites as shown in equation (12) below.

$$\lambda_{r,g,sj,t,tm}^R = 0 \qquad \forall \ g, sj, t, tm, r = CO_2$$
(12)

Equation (13) states that the net stored volume of  $H_2$  in interseasonal caverns over an annual time horizon must equal zero.

$$(\lambda_{r,g,sj,t,tm}^{S} - \lambda_{r,g,sj,t,tm}^{R}) OT_{t} = 0$$

$$\forall g, tm, r = H_{2}, sj = \text{Interseasonal cavern}$$
(13)

Equation (14) and (15) confines the heating supply in a region to be either natural gas or H<sub>2</sub> through the introduction of the binary variable,  $\alpha_{q,tm}$ . They are written as follows:

$$P_{pj,g,t,tm}^{R} \le \alpha_{g,tm} M \qquad \forall \ g,t,tm,pj = H_2 \text{boiler}$$
(14)

$$P_{pj,g,t,tm}^{R} \le (1 - \alpha_{g,tm})M$$
  

$$\forall q, t, tm, pj = \text{Natural gas boiler}$$
(15)

where M is a large positive real number. These two constraints will enable for the evaluation of regions where the usage of a particular resource is optimal in a net-zero environment. Equation (16) ensures that if  $\alpha_{g,tm}$  is non-zero, it remains so until the end of the planning horizon. Physically, this means that if a region has adopted H<sub>2</sub>, it will not be using natural gas for heating.

$$\alpha_{g,tm-1} \le \alpha_{g,tm} \qquad \forall \ g,tm \tag{16}$$

Equation (17) ensures that both onshore and offshore pipelines are separated out in order to cost them separately along with the compression requirements. It is written as follows:

$$N^{D}_{q,q',od,tm} \le ON_g K \qquad \forall \ g, g', od, tm$$
(17)

where K is any positive real number greater than or equal to 1 to constitute an upper limit on the number of pipelines of a given type.

#### **1.3** Discretisation of space

GB is discretised into 54 grid cells (polygons achieving 80 km x 80 km resolution on average), g as a compromise between accuracy and computational complexity. The centroids of these cells are subsequently used to compute the euclidean distances between one cell and another for establishing transport links<sup>1</sup>. The cells capture geographical features such as existing natural gas import locations, gas and electricity infrastructure. The total quantity of an entity in a specific cell is determined through aggregation of that entity over the entire cell area. The spatial aggregation and discretisation was performed in an open-source GIS software, QGIS 2.18<sup>2</sup> using specific tools developed for H<sub>2</sub>-CCS data processing by the authors as part of the ERA-NET ACT project "Elegancy" <sup>1</sup>. The GIS spatial discretisation in the input data processing tool was achieved using a subset of functions from within the "Resource Mapping Tool" by Cooper<sup>3,4</sup>.

## 1.4 Discretisation of time

This study considers a discrete representation of time where the temporal variations in heating demand across an annual horizon are discretised. The difficulties associated with capturing intra-day and inter-seasonal variations in demand are often described as key challenges in numerical representation of design problems. In addressing this challenge, a mathematical set, t, comprising 15 periods was

<sup>&</sup>lt;sup>1</sup>https://www.sintef.no/elegancy



Figure 1: Discrete representation of the domestic heating demand over the annual time horizon. The month of June is equated to the starting timeframe of the optimisation. Three distinct clusters of days are presented; clusters 1, 2 and 3 contain 160, 45 and 125 days respectively.

chosen so that short-term dynamics can be incorporated into the model whilst also considering seasonal variations. The resulting discrete characterisation of aggregate domestic heating demand is illustrated in Fig. 1. The specific procedure employed for formulating the discrete time representation is described in more detail as follows.

k-means clustering was used to identify 3 distinct clusters of points from nation-wide demand data. Following which, the daily variation of demand in time was discretised by minimising the error between the discrete aggregate profile and the actual data using variable length time-steps, except for cluster 2 which contains the peak demand. For cluster 2, the daily demand using the discrete profile was scaled to coincide with the peak daily demand to ensure adequacy in design. This results in an overestimate of the total heating demand on a daily basis as the discrete daily profile used in cluster 2 has a greater area than a curve traced through the actual demand data points on the peak day. Therefore, the total number of representative days in cluster 2 was scaled to ensure that the aggregate heat demand attributable to cluster 2 using discrete days is equivalent to the aggregate total of the demand data. Fig. 1 highlights the temporal demand variations over a time frame of 330 days as opposed to 365 days due to the shrinkage of representative days during scaling in cluster 2.

The effect of this choice on model outputs is discussed in Supplementary information section 3.1. Two mathematical sets are used for the representation of time, with the first (tm) used for the purposes of defining investment decisions and the second, t used to distinguish the operational decision space.

### 1.5 Input summary

- Spatio-temporal demand/ availability of all resources in the system.
- Locations of existing production, transportation and storage infrastructure.
- Techno-economic parameters and performance metrics characterising technologies CapEx, OpEx, efficiency, capacity, ramp rates, *etc.*

#### 1.6 Output summary

• Location, type, capacities of production, distribution and storage technologies.

- Production, consumption, import, flow, storage and retrieval rates of all resources in the system in all regions.
- Regional transition pathway from natural gas infrastructure to  $H_2$  at varying levels of decarbonisation.

## 1.7 Modelling assumptions

- Low pressure distribution pipelines are assumed to be available. The costs of pressure reduction stations are assumed to be negligible compared to the network components.
- Temperature variations in heating demands are not considered. Domestic and industrial boiler efficiencies were assumed to be 94% and 90% respectively for both natural gas and  $H_2^{5}$ .
- Both domestic and industrial appliance conversion costs are not explicitly accounted for in the network design. The authors assume that natural gas and H<sub>2</sub>-based appliances are comparable in both performance and total costs.
- A carbon-neutral electricity grid by 2050 is assumed as per legislation in addition to 1.5 % release in methane supply chain, amounting to 8.7  $gCO_2/MJ$  HHV of imported gas<sup>6</sup>.

## 2 Methodological innovations

In this section, a hierarchical modelling approach, specifically constructed for the assessment of model accuracy in the representation of time, is highlighted. Additionally, a multistage mathematical optimisation approach, developed for studying the transitions from existing infrastructure, is described as an alternative to multi-period model formulations.

## 2.1 Hierarchical analysis: Time

The formulation of discrete time periods could constitute a significant source of inaccuracy in the model results. A coarse representation of time is used initially as a compromise to improve computational performance. The effect of such compromises on the accuracy of model outputs is evaluated using a two-stage analytical approach to determine whether the infrastructure is appropriately sized to supply the required time-varying demands.

Stage 1 of this formulation identifies the optimal mix of technologies and infrastructure required to achieve the least cost design for a target reduction in emissions, using a collection of 15 time periods. Stage 2 initially assigns the integer variables (*i.e.*, quantity of a specific production, transportation and storage technology installed in a given location) with the respective outputs obtained from stage 1. Upon which, the temporal resolution of the operational time is improved by representing the same data using a greater number of discrete time periods. The subsequent problem is re-optimised to determine whether the resulting network has been over/under-sized. If the model instance from stage 2 is proven to be infeasible, then the network has been under-sized and thus, incapable of meeting the heating demands. In this case, ancillary investments are required to support the additional capacity requirements. The amount of additional investment required is defined via the relaxation of the initial assignments, and further incorporation of integer outputs as the lower bound for the realisation of the investment variables in the subsequent optimisation.

The modelling workflow is graphically depicted in Fig. 2. The inequality constraint in stage 2 reduces the size of the feasible set of solutions, enabling the model instance to be solved at a higher temporal resolution. Consequently, an optimal solution achieved via this method will clearly detail the additional infrastructure requirements to ensure technical feasibility and operability. When stage 2 presents a feasible solution, it is possible to identify the extent to which the infrastructure is over-sized. It is worth noting that this modelling approach can be applied sequentially, in an iterative manner with



Figure 2: A schematic of the workflow used for evaluating the effect of temporal uncertainties where y and x denote the set of investment and operational decision variables respectively. The numerical subscript refers to the realisation of these variables within the corresponding stage and the asterix is used to denote the optimal values for the decision variables.

increasingly higher temporal resolutions to eventually reflect that of the original data set. The main disadvantage associated with this approach is that solution optimality is only guaranteed at the highest level of the hierarchy (*i.e.*, stage 1) and improved representations of time result in solutions which depart from mathematically guaranteed optimum to those that are practicable in design. The authors note great value in the development of mathematical frameworks and techniques that are capable of optimising the trade-offs between high spatial and temporal resolutions in planning problems.

## 2.2 Transition pathway

A multi-stage optimisation model variant was developed to study the transition pathway from natural gas infrastructure to H<sub>2</sub>. Typically, multi-period model formulations are used for these purposes but the presence of large groupings of integer decisions in multi-vector design problems could present severe computational limitations for problems described over a long-term planning horizon. Instead, it is possible to use snapshot model formulations (*i.e.*, single investment time horizon) to define a phased deployment pathway in instances where the inherent model components are not susceptible to significant change in the long-term. There is inherent uncertainty in the assumption of prices for resources such as natural gas, electricity and biomass over time. However, these elements are not forecasted and exogenously imposed onto the model based on authors' assumptions. The resulting model formulation used in this paper utilises a snapshot solution with an iterative component which defines the deployment rate of technologies as a function of the target reduction in  $CO_2$  emissions intensity. These model outputs are subsequently matched with a time horizon through the satisfaction of  $CO_2$ emissions intensities at respective time points. In the pathway analysis, stage 1 of the formulation is identical to that in 2.1. Subsequently, the emissions intensity constraint is updated. Upon which, the model is re-optimised using the initial solution from stage 1 as the upper bound for the design variables, y in the next stage. This significantly reduces the solution space and limits the technological options to only those present in stage 1. This specific solution procedure is depicted in Fig. 3.

b denotes a set of binary decisions which determine whether a region is supplied with natural gas or H<sub>2</sub>. In particular, b is indexed via the set of spatial cells, g and b = 1 indicates that a region is converted to H<sub>2</sub>, whereas b = 0 implies the continued usage of natural gas. The key factor influencing



Figure 3: A schematic overview of the workflow in defining spatially resolved outputs at 5 year time intervals through a snapshot model solution and subsequent processing as opposed to using a computationally expensive multi-period formulation.

the resolution of the transition pathway in this model instance is the cardinality of the set, tm. In particular, the analyses presented in this paper assume 7 discrete intervals of uniform size with the resulting set spanning across a 35 year time horizon. Hence, a uniform interval length of 5 years is assumed per element of tm. Thus, under the assumption that  $tm = \{1, 2, ..., 7\}$ , the solution procedure shown graphically in Fig. 3 terminates when  $tm = T_{end} = 7$ . The key benefit of this modelling approach lies in the possibility for obtaining deployment pathways over shorter time intervals than what is typically achievable in similar multi-period planning models. The availability of finely resolved model outputs are invaluable for the contrivement of long-term infrastructural plans.

#### 2.3 Statistical analysis

Key statistical parameters are computed to distinguish the importance of input parameters whilst ensuring reliable results. They are comprised of the pearson correlation coefficient (r), adjusted coefficient of determination ( $\bar{R}^2$ ), standardised regression coefficient ( $\beta$ ), Variance Inflation Factor (VIF) and the F-ratio. The r value denotes the strength of the linear correlation between model inputs,  $Z_i$ and outputs, Y where Z is the vector of randomly generated input values for uncertain parameter *i* and Y is the vector of model outputs. The F-ratio is the ratio of the variance explained by the regression model relative to its unexplained residuals. This parameter is compared with the F-statistic to understand whether the addition of another input in the regression model results in statistically significant improvements in its predictive ability. Equations for  $\bar{R}^2$ ,  $\beta$  and the VIF are defined as:

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1}$$
(1)

$$\beta_i = RC_i \frac{\sigma_{z,i}}{\sigma_y} \qquad \forall \ i \tag{2}$$

$$VIF_i = \frac{1}{1 - R_i^2} \qquad \forall \ i \tag{3}$$

where  $\mathbb{R}^2$  is the coefficient of determination, n is the sample size, p is the number of predictive inputs,  $RC_i$  is the regression/ partial coefficient for uncertain parameter i in the regression model,  $\sigma_{z,i}$  and  $\sigma_y$ represent the standard deviations for  $Z_i$  and Y.  $\overline{R}^2$  adjusts  $\mathbb{R}^2$  based on the number of independent variables and the sample size. The  $\beta_i$  coefficient describes the proportionate impacts of one standard deviation of change in the uncertain input parameter i on the standardised model output. Therefore,  $\beta_i$  values can be directly compared and allows for the relative importance of the inputs to be deduced from the regression model. Finally, the  $VIF_i$  measures multicollinearity between uncertain model input parameters. A low degree of multicollinearity is observable at lower  $VIF_i$  values, whereby the resulting variance can be attributed to independent input parameters as opposed to a collection of inputs to improve accuracy of estimation. A stepwise estimation procedure based on Hair et al.<sup>7</sup> is used to build the regression model.

## 3 Supplementary results

#### 3.1 Representation of time

The TAC was computed via independent optimisations of the model at pre-defined intervals of 5% reduction in  $CO_2$  emissions intensity. The resulting solutions were studied to determine whether the designs suffice or require additional investments using the two-stage hierarchical modelling approach. The discernible differences in total costs between stages 1 and 2 are plotted in Fig. 4. The difference in TAC between 15 and 60 time periods is expressed through the definition of a new variable  $TAC_{diff}$  as follows:

$$TAC_{diff} = (TAC_{60} - TAC_{15})/TAC_{15}$$
(1)

where  $TAC_{15}$  and  $TAC_{60}$  refer to the total annualised costs of heat supply using the model formulation with 15 and 60 time periods respectively. Fig. 4 demonstrates that  $TAC_{diff}$  equates to zero only at the origin where the  $TAC_{60}$  and  $TAC_{15}$  are necessarily indistinguishable. Natural gas is continually



Figure 4: A scatterplot of the differences in TAC between 60 and 15 time periods expressed as a percentage plotted against maximum achievable  $CO_2$  removal through independent cost-optimal designs.

supplied without any additional infrastructural investments when there is no effective  $CO_2$  emissions reduction constraint. Thus, total costs of heat supply remain identical in both stages irrespective of the temporal resolution. A point to emphasise is the prominence of the data in the positive real space indicating that using a set of 15 time periods results in an underestimate of the required investments in this instance. Thus, the initial representation of time results in solutions that are infeasible in practice. The largest error in TAC across all optimisations was within 3% of the optimum solution, occuring at the highest CO<sub>2</sub> removal targets. Although the network has been undersized in all instances, minimal capacity additions are required for ensuring operability of design. Thus, the representation of operational time was deemed appropriate for this design problem.

## 3.2 Peak day operating mix for H<sub>2</sub> production

Fig. 5 displays the operating capacity during peak supply in the absence of  $H_2$  storage. ATRs, SMRs and BG with CCS are operational throughout the course of the day with elevated production rates at times of peak demand. In contrast, WE units are only operational for the satisfaction of peak demand due to their higher OpEx requirements. There is little commercial experience and academic literature surrounding the flexible operation of SMRs, in part due to its limited use in applications with temporal demand variability. Thus, operational flexibility criteria must be developed as part of future research. At present, large scale deployment of the WE process should be recommended for applications with significant temporal variabilities in the absence of sufficient  $H_2$  storage. The resulting power requirements for WE might necessitate capacity expansions within the electricity transmission grid especially when peak demands for the given application and power use coincide, amplifying the magnitude of associated investments.



Figure 5: Operating mix during the peak day in the absence of  $H_2$  storage. The importance of the WE process for supplying peak demands is shown here.

## 3.3 Statistical analysis of the regression models



Figure 6: Scatterplots to investigate the correlative behaviour between resource prices, cavern costs and the total annualised cost, TAC using 10,000 samples. Gas price appears to have the strongest correlation with TAC whereas the electricity price shows the weakest, if any. Both biomass price and cavern cost seems to have a strong influence on TAC.

4 scatterplots were generated using randomly generated samples for each input parameter and its corresponding output as depicted in Fig. 6. The quadrants were generated using the mean for the input and output TAC. Observations from the scatterplots highlight a stronger correlation between gas price and TAC in comparison to other resources. In contrast, the electricity price seems to have little effect on the TAC as illustrated by an approximately equal occupancy of data points within each quadrant. From the figure, it is not entirely clear whether the TAC is more sensitive to variations in biomass price or cavern cost.

Multicollinearity between uncertain input parameters was analysed using the  $VIF_i$  where i is an element from the set, {gas price, biomass price, electricity price, cavern cost}. All VIF values were determined to be close to unity within 3 decimal places, implying a lack of correlation amongst the input parameters. However, there may be underlying correlative effects describing the electricity price as a function of the gas price rather than as a uniform distribution across the uncertainty space. Therefore, dependencies between these inputs could potentially distort its impact on TAC. Nevertheless, regression coefficients and goodness-of-fit statistics are combined to determine the overall contribution from potentially correlated inputs as shown in Table 4. Within the table, model A includes gas price, biomass price and cavern costs as the explanatory variables whereas model B includes electricity price in addition to the aforementioned variables. Therefore model A contains three degrees of freedom (D.O.F) and model B contains four.

Table 4: Standardised regression coefficients and statistical analysis for multivariate regression models to scrutinise the inclusion of electricity price as an explanatory variable.

	Model A	Model B
D.O.F	3	4
$\bar{R}^2$	0.984	0.987
F-Ratio	$2.00 \times 10^{5}$	$1.85{ imes}10^5$
$\beta_{gas \ price}$	0.709	0.708
$\beta_{biomass\ price}$	0.513	0.514
$\beta_{cavern\ cost}$	0.465	0.464
$\beta_{electricity \ price}$	-	0.055

The differences between  $\bar{R}^2$  and the F-ratio must be noted when comparing models, A and B. In particular, the F-ratio decreases upon the inclusion of electricity price as a regression variable whilst displaying a negligible increase in the  $\bar{R}^2$  value by 0.003. Thus, model B was rejected, and model A was confirmed as the most appropriate representation for a multivariate regression model. The t-statistic was used to ascribe a confidence interval to the regression coefficients in order to confirm that the estimated regression coefficients are statistically significant. The t-statistic is evaluated using the estimated regression coefficient and the standard error of estimation. Across all  $\beta$  coefficients, the maximum error is within  $\pm 0.6\%$  of the standardised value which leads to the conclusion that all of the estimated coefficients are statistically significant since the confidence intervals do not overlap with zero.

## **3.4** Impact of grid CO<sub>2</sub> intensities

The impact of electricity and natural gas grid  $CO_2$  intensities on total costs and energy consumption was evaluated. Complex system wide trade-offs arise from the presence of upstream supply chain emissions. The extent of their influence on total energy requirement and TAC for complete decarbonisation of the heating supply is illustrated via Fig. 7.

The percentage increases in total energy consumption were computed as a function of grid  $CO_2$  intensities, measured relative to the energy requirements that arise from a carbon neutral electricity and natural gas grid. The figure affirms the relative importance of reducing upstream methane supply



Figure 7: An illustration of the percentage increases in primary energy consumption and TAC compared to the requirements at grid CO<sub>2</sub> intensities of 0 g/kWh.

chain emissions. Note that the  $CO_2$  footprint of the natural gas grid has a greater influence on the total energy consumption than that of the electricity grid. At higher  $CO_2$  intensities, the requirement for negative emissions should increase. Thus, resulting in a greater deployment of BG with CCS, which increases the overall energy consumption due to its lower overall energy efficiency. This leads to the expectation of higher increase in energy requirements than that observed at higher  $CO_2$  intensities. This peculiarity arises from the fact that BG with CCS is deployed to its optimal extent at net zero emissions targets. It appears economically optimal to combine BG with CCS for H<sub>2</sub> supply in certain regions whilst continuing the supply of natural gas in other regions. A direct implication of this observation is that a greater deployment of H<sub>2</sub> is necessitated by higher grid  $CO_2$  intensities, thereby eventually converting all of the natural gas supply to H<sub>2</sub>.

Similarly, the effect of grid  $CO_2$  intensities on the total costs for decarbonisation is also illustrated in Fig. 7. The resulting surface illustrates the potential for the *TAC* to increase by over 15%, displaying a stronger correlation with increasing  $CO_2$  intensities. The extent of this variability provides a clear incentive for reducing the upstream methane supply chain emissions. This is a crucial point to consolidate as few studies have extensively focused on the reduction of emissions from gas grids in comparison to the electricity grid. There is little interdependence between the electricity grid and the H<sub>2</sub> supply system in the absence of significant WE deployment. Naturally, the relative importance of the decarbonisation of the electricity sector is magnified when the WE process amasses a greater share of the production mix.

#### **3.5** Impact of biomass availability and CO<sub>2</sub> intensity

The impact of biomass availability and  $CO_2$  intensity on optimal value chain design must be understood. Note the distinction between biomass supply  $CO_2$  intensity and the embodied emissions that are released during the gasification and conversion of the feedstock. The former explicitly refers to the direct and indirect greenhouse gas emissions across the biomass supply chain with its origins from sourcing and transportation before use. Fig. 8 illustrates the effect of biomass availability and its



Figure 8: An illustration of the percentage increase in total annualised system costs compared to the minimum cost requirements at a supply  $CO_2$  intensity of 0 g/kWh with varying potential availability of indigenous biomass.

supply chain intensity on TAC for decarbonisation of the heating supply. As expected, the TAC of decarbonisation is highly influenced by both the  $CO_2$  footprint of the biomass supply chain and the total availability of biomass. In particular, significant economic penalties are associated with a constrained supply of biomass. At least 150 TWh of biomass is required for a complete decarbonisation of the GB H<sub>2</sub> supply network at least cost. In addition, the sustainable sourcing of biomass is of paramount importance for the reduction of TAC, as evidenced by a cost increase as high as 30% as shown. These findings have major implications on the extent of mobilisation of land for the sourcing of biomass and its transportation means, which must be considered in the long-term planning of H<sub>2</sub> infrastructure.

## References

- O. Akgul, A. Zamboni, F. Bezzo, N. Shah and L. G. Papageorgiou, *Ind. Eng. Chem. Res*, 2011, 50, 4927–4938.
- [2] QGIS, QGIS The Leading Open Source Desktop GIS, 2017, https://www.qgis.org/en/site/ about/index.html.
- [3] Nathanial Cooper, *GitHub nathanialcooper/Resource-Mapping-Tool*, 2018, https://github.com/nathanialcooper/Resource-Mapping-Tool.
- [4] N. Cooper, A. Panteli and N. Shah, Applied Energy, 2019, 253, 113526.
- [5] M. Chaudry, M. Abeysekera, S. H. R. Hosseini, N. Jenkins and J. Wu, *Energy Policy*, 2015, 87, 623–640.
- [6] P. Balcombe, K. Anderson, J. Speirs, N. Brandon and A. Hawkes, Sustainable Gas Institute, 2015.

[7] J. Hair, W. Black, B. Babin and R. Anderson, *Multivariate data analysis*, Pearson Education Limited, 7th edn, 2006.