MAC-MIGS CDT

PhD in Mathematics

MAXWELL INSTITUTE FOR MATHEMATICAL SCIENCES

Optimal Infrastructure Planning for the Removal of CO_2 from the Atmosphere.

Authors:
Lucas Beerens
Liam Llamazares
Christopher A Oldnall

Supervisors:
Prof. Miguel Anjos
Dr. Hannah Chalmers

December 13, 2021

THE UNIVERSITY OF EDINBURGH

Abstract

In this paper we explore the usage of stochastic optimization in deploying an infrastructure built of pipes and ships within the UK in order to transport CO_2 from clusters (sources) to sequestration sites in parts of the Northern Sea and East Irish Sea (sinks). After visiting some background of the problem, this paper looks to implement a multistage stochastic linear framework with a nodal formulation. The implementation will lead to a model which can show the most efficient layout of the infrastructure to transport the most CO_2 based on a stochastic variable of unknown CO_2 capture from each cluster.

For model files and assets please see:
`www.github.com/LucasBeerens/stochastic-optimisation-co2-transport`

Contents

Acronyms

CCC Committee on Climate Change. 1

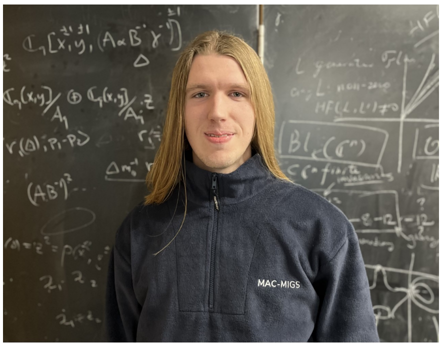
CCUS carbon capture usage and storage. 1, 3, 5

Acknowledgements

With thanks to **Professor Miguel Anjos** and **Dr Hannah Chalmers**, without whose guidance, enthusiasm, and passion in the field of optimization and CO_2 removal this project would not have come to light.

Author Contributions

Lucas Beerens
l.beerens@sms.ed.ac.uk

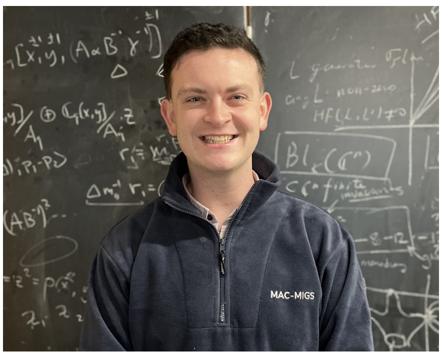


Throughout the project, Lucas was responsible for building the link between the deterministic and stochastic programs, including the key solution of the nodal formation. He then took the lead in implementing the model within FICO Xpress.

Liam Llamazares
l.s.llamazares-elias@sms.ed.ac.uk

Liam's role within the project was initially to take the lead on building the model which would go into the programe. This led to key research, scanning the already existing literature for relevant pricing and model information.

Christopher A Oldnall
chris.oldnall@ed.ac.uk



Beginning with background and political research, Chris' role in the project was to collate the information that was needed to understand the usage of the model and data. This then led to him taking a lead on the copywriting side of the project.

Part I
Introduction

> "Naturally, we can propose many sophisticated algorithms and a theory but the final test of a theory is its capacity to solve the problems which originated it."
>
> George Dantzig [4]

The everyday world that we live in is encapsulated by the decisions that we as humans make. As a result, vast problems are created that need to be solved, optimized, and interpreted. Fortunately, there exist many different mathematical techniques in approaching these.

Consider the situation where you want to check how to get from A to B in the shortest time, using roads and public walkways. Traditionally we would call this a **Deterministic Optimisation** problem. A number of algorithms could be run and you are left with a solution to the route you should take. Now add into that scenario that you do not know where B actually is, simply a rough estimate, and that the only maps you have access to are from the 1970s. All of a sudden you cannot necessarily find the one set path, B could be further out than expected or paths could no longer exist. What we have imagined here is deemed to require **Stochastic Optimisation**. Throughout this paper, we will be exploring how to bring together stochastic optimization methods to solve one of the biggest problems being tackled in the climate change sector, carbon capture usage and storage (CCUS).

The Problem:

In 2018, the Committee on Climate Change (CCC) for the UK identified that CCUS was vital to achieving a significant drop in CO_2 emissions [3]. The CCC report detailed 6 industrial centres which a potential project could take into account. We will refer to these sites as '**(industrial) clusters**'. Later on, it has been identified that there are four main sections under the ocean beds of the Northern and Irish seas which each have several sections where CO_2 could be stored. For us these are '**sequestration sites**'. The background of how we will model these will be addressed in Part II.

> **Research Question:**
> Can we build a model which optimally plans infrastructure for the implementation of CCUS across the UK, without knowing the exact amount of CO_2 we are able to capture from a cluster?

In our model, a set of decisions will be made around the building of a network of pipes and ships as well as infrastructure for sequestration within given budgets of capital and operational costs (later to be referred to as **CAPEX** and **OPEX** budgets respectively) in order to transport the CO_2 from clusters to sequestration sites. Pipes and ships are selected as the available pieces in our infrastructure as both are already highly obtainable and accepted by those working with CCUS.

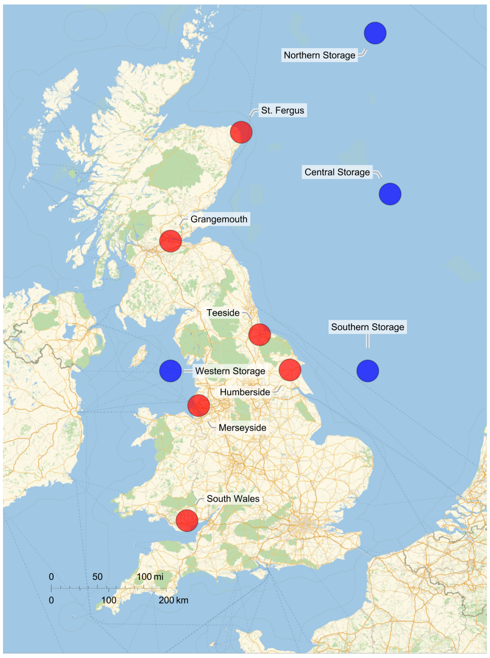


Figure 1: Map of the UK showing clusters and sites. (Image from: Mathematica)

Mathematical Rigour:

This paper will also aim to explain the use of stochastic optimization utilized in our problem. In particular, we will be looking at how our model is built around the basis of using a **multistage stochastic linear program** of the form,

$$\max\{b_0 \cdot x_0 + \mathbb{E}_{\tilde{\xi}_1|\xi_0}[Q_1(x_0,\xi_1)] : x_0 \in S_0\},$$

which takes a recursive **benefit-to-go** function Q_t. These terms will be explained in Part III. The model we have created, and will explore in this paper, tries to take into account most of the constraints of the real-life situation. Some constraints which are vital to the situation are: the situation proposed depends on a pre-designated budget, once a sequestration site is full we cannot add any CO_2 to it, the ships and pipes transport an amount of CO_2 which is between 0 and maximum capacity.

The Correct Choice:

The key idea which sits behind stochastic optimization is that at the end of running the model we will be provided with a decision that maximizes the objective function, in our case the amount of CO_2 stored, which is calculated via an expectation across all possible scenarios. Part IV will engage in a discussion of how we implement the optimization programme in the language Mosel. Some functions to conduct a future sensitivity analysis will also be suggested.

For ease of exploration of this paper we once again note here the following structure:

- Part II explores the background of the model,
- Part III presents the reader with the mathematical model and constraints,
- Part IV runs our model for some data and looks to where analysis could be performed.
- Part V will draw some conclusions for our project.

Part II
Background of the Model

"Everything is physics and math."

Katherine Johnson

The UK government has expressed the goal of deploying CCUS from 2025 [5]. To recap, our problem lies around looking to optimize the amount of CO_2 which is transported from clusters to sequestration sites. The model we will build in this paper considers that the infrastructure for the CCUS will consist of pipes and ships. In a real world implementation, the CO_2 would take the following journey through the infrastructure,

1. CO_2 is captured at each cluster.
2. In the case where the CO_2 is transported through pipelines it is pressurized, if the CO_2 is transported through ships it is also liquefied (cooling and compressing in order to increase density for cost-effective transport) at a liquefaction plant.
3. Temporarily CO_2 is then stored in liquid form in tanks. This is necessary as CO_2 is produced continuously but for ships, they will arrive in discrete runs.

If ships to transport the CO_2 then it would be loaded onto a carrying ship and take one of the following paths.

- **Port to Port Shipping:** The CO_2 is unloaded onshore in liquid form to temporary storage tanks. Then, it is pumped at heated and finally transferred via pipelines to a long-term storage site. This is the plan established by the Norwegian CCUS and it is estimated that 4Mt of CO_2 could be stored per year (for reference the total CO_2 production of Norway is 43Mt per year).
- **Port to Storage Shipping 1:** The CO_2 is pumped and heated on the ship and transferred to the injection well of an offshore storage site.
- **Port to Storage Shipping 2:** The CO_2 is transferred in liquid form to an offshore platform where it is stored temporarily and then pumped and heated and then stored in an offshore storage site.

Alternatively, if pipes are used for transportation, then pipeline from shore to an offshore storage site is built through which to send the CO_2 .

In all the situations above it may seem that a solution is obvious to the reader, simply link each cluster with the nearest site which could hold the capacity. This would be the deterministic optimization situation. However as earlier addressed we are in a stochastic setting. This leads us to a reconsideration, what happens if we don't know how much CO_2 is captured or emitted from a certain site? In this case the solutions is not so clear and it is necessary to develop a more sophisticated model.

1 Deterministic vs Stochastic Optimization

In any of the cases presented above in terms of CO_2 shipping, one would be able to set up a deterministic program which would tell us which ships and pipes should be placed where and how much CO_2 would be stored in each sequestration site. This deterministic program would return an optimal setup of our infrastructure, denoted by x_0^* as below.

Definition 1: Deterministic Optimization Formulation

$$x_0^* = \arg\max_x \left\{ c \cdot x \middle| Ax \leq b \right\},$$

where c are the objective coefficients, and $Ax \leq b$ are the system of linear constraints.

Of course this set-up would be ideal if we could measure the exact storage capacity of a sequestration site and the amount of CO_2 that a cluster produces. Unfortunately we cannot and therefore we are left in the stochastic situation whereby we have to take the average of a bunch of scenarios and present an individual with an expected result based on the chances of different scenarios occurring. A stochastic optimization formulation is presented below.

Definition 2: Static Stochastic Optimization Formulation

$$x_0^* = \arg\max_x \left\{ \mathbb{E}[F(x,\xi)] \middle| x \in X \subseteq \mathbb{R}^n \right\},$$

where X is some decision vector, ξ is a stochastic process and F is a function which progresses the process.

We see that indeed now whilst we are still taking the minimum of the problem, the objective coefficients and our optimiser, x, have been replaced with the expectation of a function that progresses the process. Later on in this paper, we will explore the form of the function more thoroughly. It should hopefully by now be clear to the reader why we are in the stochastic setting for our optimization, but should an individual want to read more into this the work of Ahmed [1] may be useful for building the understanding link.

2 Clusters, Intermediate Storage and Sequestration Sites

For the purposes of this project, we will work on the assumption that the unknown size of the sequestration sites will be covered by us generalising these to four areas. That is to say, we will be looking at four parts of the sea which each contains dozens of individual sequestration sites under the ocean bed and our model will not reach the situation where they become full or unable to take a load of CO_2 required. Of course, a development of this model could be to incorporate all the ‘sink’ nodes that the individual sites present and investigate how stochastic optimization would amend the results to an exact site.

In our problem, the stochastic nature of the CO_2 produced at the clusters will be preserved and our problem will look to optimize while operating under this unknown. Typically in a true model, we would split the journey of the CO_2 into three different types of nodes:

1. **Clusters (CL):** These are the initial sites from which the CO_2 is collected.
2. **Intermediate Storage (IS):** Here the CO_2 is either stored in storage tanks before it can be loaded onto ships or potentially gets sent through an alternative pipe to a site.
3. **Sequestration Sites (SQ):** The endpoint of the CO_2 where it is deposited into the under-ocean storage parts.

An important decision we have made for the purposes of this report is that we will absorb all potential ‘IS’ nodes into the group of ‘CL’ nodes. That is to say, we will simply assume that CO_2 only flows directly from a cluster to another cluster or sequestration site. The reasoning behind this decision is that the we consider large time steps and will simply include any additional costs within the price of transportation. That said, perhaps with more time one could investigate the effects of including intermediate storage in the model.

3 Phased Rollout

The UK government has recently made decisions around the phased rollout of CCUS. It has been decided that the Teeside cluster will be the first to implement the carbon capture infrastructure. Following this, the Merseyside, Humberside and St Fergus clusters will be next. After this, the rest of the clusters will be rolled out. We will incorporate this information into our model through the information that we feed it. Whilst in the next part we will build a model which considers the entire CO_2 emissions of a cluster to be stochastic, we will then constrain our model with expert data to only consider the amount we can capture as stochastic as ultimately this limits our final goal more.

In our model we will incorporate a finite set of time steps. This is so that we can consider different budgets at different time steps with which to build infrastructure as would occur in a real world implementation. In part we also do this as the model may be tempted to build all infrastructure immediately in time step 1. This unfortunately in the real world isn’t able to happen due to some areas not being prepared for a variety of reasons and hence will not be awarded funding.

Part III
Dynamic Modelling

"It is better to solve one problem five different ways than to solve five problems one way."

George Polya

Having explored the background to our problem in the previous section, we will now look to explore how we will go about creating a dynamic model. As mentioned previously, we will be implementing a stochastic optimization. This will affect the way our objective function is constructed and how our decision variables are constrained. The structure of the parameters and the constraints will look akin to those of a deterministic programming setup. We develop our model in Section 4 before specifying the values of its parameters in Section 5.

4 The Model

Our model aims to optimize the amount of CO_2 sequestered given a certain budget. Within this we however wish to consider that the following may vary:

1. CO_2 emissions,
2. Money invested.

The model itself is built up to consider CO_2 emissions as a stochastic variable, whilst the money invested will be varied manually later on when testing the model. Below we present the framework of multistage stochastic linear programming (MSSLP). We will be modelling our CO_2 capture process using an MSSLP, as the decisions around infrastructure implementation will be made in discrete time steps.

Definition 3: Multistage Stochastic Linear Programming (MSSLP)

For some sequence of decision vectors through time, $\xi_t = (\xi_1, \xi_2, ..., \xi_T)$ we have that the optimal solution x_0^* is given by,

$$x_0^* = \operatorname*{arg\,max}_{x_0 \in S_0} \left\{ b_0 \cdot x_0 + \mathbb{E}_{\tilde{\xi}_1 | \xi_0} \left[Q_1 \left(x_0, \xi_1 \right) \right] \right\}$$

where we define the *benefit-to-go* function Q_t recursively by:

$$Q_t \left(x_{t-1}, \xi_t \right) = \max_{x_t \in S_t(x_1, \ldots, x_{t-1}, \xi_t)} \left\{ b_t \cdot x_t + \mathbb{E}_{\tilde{\xi}_{t+1} | \xi_t} \left[Q_{t+1} \left(x_t, \xi_{t+1} \right) \right] \right\}$$

and $Q_{T+1} = 0$.

The interpretation of the above formulation is as follows:

- ξ_t is the stochastic information (or realization) that we learn in each step.

- x_0 is the first stage decision and x_t is the decision (or policy) at time t. It is important to note that, while x_0 is well defined independently of any realizations ξ_t or further decisions x_t, the decision x_t are themselves dependant on previous decisions and realizations. Note for example that we require that x_t be in the constraint set $S_t(x_1, ..., x_{t-1}, \xi_t)$. The fact that x_t depends on the realizations of some random processes implies that it is a priori impossible to define what an optimal solution $(x_0, ..., x_T)$ is or what it looks like. We will discuss later in 6.1 how we deal with this problem in a way that satisfactorily captures not only what an optimal initial decision x_0 is but also how we can expect our decisions to evolve. Further, we note that in the standard formulation the parameters b_t and the constraint sets S_t only depend on the values of x_t and ξ_t. In our case, it is also necessary to incorporate dependence on previous decisions $(x_0, ..., x_{t-1})$. We will elaborate later on why this is the case.
- b_t is the benefit of implementing a decision x_t. In a general case, this could depend on ξ_t but in our case, this will not occur as we will see later.
- Our time steps will be 5 years 'jumps' and go through decisions that need to be made after 2025 through to 2050.
- The benefit-to-go function is the single evaluation of the stochastic optimization problem one-time step in the future.

Let us consider why it is useful for us to formulate our problem as an MSSLP. It should by now be clear to the reader that decisions around infrastructure in the system will be implemented in time steps. This demands from our set-up that we should be able to inform later calculations of the program based upon earlier decisions. Furthermore, it is important to recognize that due to the stochasticity of the CO_2 release we also have the situation where the random variable will update in each time step.

4.1 Glossary

Below we define the sets and parameters that will be employed in the model. It is recommended that the reader becomes familiar with the notation so that when it comes to considering the constraints on the model it is easier to understand the mathematical meaning. Here we note that we will use super indices. Consider a product space of the form,

$$X = A^a \times B^b,$$

with $a, b \in \mathbb{N}$. Then, given $x \in X$ we use super-indices to denote whether we are considering the component of x in A^a or in B^b, that is x is has as components:

$$x^{(i)}_{j,k} \quad i = 1, 2; \quad j = 1, ..., a; \quad k = 1, ..., b.$$

4.1.1 Sets

- CL is the set of industrial clusters where CO_2 is captured. This set contains 6 elements as marked red in Figure 1. We index this set by the sub-index cl.
- SQ denotes sequestration sites where CO_2 is placed under the ocean bed. This set contains 4 elements as marked blue in Figure 1. We index this set by the sub-index sq.
- Due to considering transport from elements of CL to other items in CL and in SQ we introduce the union $CL \cup SQ$. This set will be sub-indexed by cs.

- S_p are the types of pipes that can be built. We will consider pipes of diameter $10, 20, 25, 30$ and 35 inches as were studied by MIT [6]. We index this set by the sub-index sp.
- S_{sh} are the types of ships that can be built. We take these to be ships of capacity $10000, 20000, 30000$ tonnes of capacity as were studied by BEISD [2]. We index this set by the sub-index sh.
- K contains the kinds of ways in which CO_2 can be captured. We will consider the 3 levels of CO_2 capture: cheap, medium and expensive. Here cheap and medium will be coming from different types of power-plant capture whilst the expensive will be direct air capture. We index this set by k.
- $\{0, 1, ..., T\}$ are the time steps of our model. This set will be indexed by either t or s where appropriate.

Furthermore, to simplify many expressions we will write expressions such as $\forall cl$ to signify $\forall cl \in CL$. An analogous notation is used for all other indexes and their respective domains.

4.1.2 Parameters

- $a = |CL|$ is the number of industrial clusters (sources).
- $b = |SQ|$ is the number of sequestration sites (sinks).
- $\lambda_{cc} \in \mathbb{R}_+$ is the cost of increasing the capture rate of CO_2 of an industrial cluster by 1 tonne.
- $\lambda_{oc} \in \mathbb{R}_+$ is the operating cost associated with the capture rate of 1 tonne of CO_2 .
- $sc \in \mathbb{R}_+^{|S_{sh}|}$ is a vector that tells us the maximum amount of CO_2 each type of ship can transport over the entire time period.
- $pc \in \mathbb{R}_+^{|S_p|}$ is a vector that tells us the maximum amount of CO_2 each kind of pipe can transport over the entire time period. We consider that this is independent of the locations between which the pipes travel.
- $scc \in \mathbb{R}_+^{|S_{sh}|}$ contains the information on the capital cost of building a ship in S_{sh}. We note that this cost is independent of the location between which ships travel.
- $soc \in \mathbb{R}_+^{a\times(a+b)\times|S_{sh}|}$ tells us the operating cost per ton of CO_2 transported. $soc_{cl,cs,sh}$ is the operational cost of transporting a ton of CO_2 $1km$ between a cluster cl and a cluster or sequestration site cs.
- $pcc, poc \in \mathbb{R}_+^{a\times(a+b)\times|S_p|}$ tell us the capital cost and operating cost of pipes. More explicitly we have $pcc_{cl,cs,sp}$ and $poc_{cl,cs,sp}$ are respectively, the cost of building pipeline of type $s_p \in S_p$ from $cl \in CL$ to $cs \in CL \cup SQ$, and the cost of sending 1 tonne of CO_2 through this pipeline.
- $qcc, qoc \in \mathbb{R}_+^{b}$ are respectively the capital and operating cost of sequestration material for $1tCO_2$.
- $MC_t \in \mathbb{R}_+$ is the CAPEX budget at time t.
- $MO_t \in \mathbb{R}_+$ is the OPEX at time t.
- $NO \in \mathbb{R}_+$ is the maximum operational budget we allow at the end of the simulation.
- $CA \in \mathbb{R}_+^{a}$ is the maximum capacity of a sequestration site.

4.2 Decision Vector

Having introduces the sets and variables we will be using, it remains to see how these are implemented to the MSSLP definition for our model. In SO the decision is the part of the model that is determined in order to maximize the benefit, in our case the amount of CO_2 produced. We set the decision at stage t to be

$$x_t = (cd_t, sd_t, pd_t, qd_t).$$

Where the components cd_t, sd_t, pd_t, qd_y respectively determine how much CO_2 we choose to capture, in addition to the infrastructure of the ships, pipes and sequestration at stage t. We now elucidate the exact structure of each of these components.

4.2.1 Capture decision

The capture decision is

$$cd_t \in \mathbb{R}_+^{a\cdot|K|} \times \mathbb{R}_+^{a\cdot|K|},$$

where $cd^{(1)}_{t,cl,k}$ indicates the capacity for CO_2 capture of type k that we add at stage t in the cluster cl and $cd^{(2)}_{t,cl,k}$ is the amount of CO_2 we capture at cl at time step t. This last element of the capture decision is motivated by the fact that it will be necessary to adjust how much CO_2 is captured based on the quantity of CO_2 that each cluster produces at time t as well as other factors such as the transport infrastructure that has been built. For example, in the case where the rate of CO_2 production of a cluster went below the capacity of our infrastructure, we would now use this CO_2 production as a bound for how much we actually capture as opposed to the amount of CO_2 our infrastructure says we could capture. Similarly, we would not capture more CO_2 than we could transport to sequestration sites.

4.2.2 Shipping decision

The shipping decision is

$$sd_t \in \mathbb{N}^{|S_{sh}|} \times \mathbb{N}^{a\times(a+b)\times|S_{sh}|} \times \mathbb{R}^{a\times(a+b)\times|S_{sh}|}$$

Where each component has the following meaning.

- The first component $\mathbf{sd}^{(1)}_t \in \mathbb{N}^{|S_{sh}|}$ indicates the number of each kind of ship that are built at time step t. Due to the time, it takes to build a ship we consider that these ships will only become available for transport at stage $t+1$.
- At each time step t each of the kinds of ships that have been built at stages $1, \ldots, t-1$ are assigned a port where they will be based and a cluster or sequestration site to which they will transport CO_2 (transport of CO_2 between clusters could prove advantageous in the case where they are sufficiently nearby and one of the clusters is connected to a sequestration location) during the current stage t (both these assignation may change as a ship may be required at different ports as CO_2 production varies). All this information is encoded in the component $\mathbf{sd}^{(2)}_t \in \mathbb{N}^{a\times(a+b)\times|S_{sh}|}$. Namely, $sd^{(2)}_{t,cl,cs,sh}$ is the number of ships of type sh that are transporting from cl to cs at time t.
- The third component $\mathbf{sd}^{(3)}_t \in \mathbb{R}^{a\times(a+b)\times|S_{sh}|}$ indicates the amount of CO_2 which ships transport between a given capture location and cluster or sequestration site at time t. This element of the decision is necessary as it may be necessary to adjust how much CO_2 is

transported based on the quantity of CO_2 captured and made available for transportation. In this case, we consider that the change in time t instantly affects the amount of CO_2 shipped in this step (and not the next one).

4.2.3 Piping decision

The piping decision presents a similar structure to that of the shipping decision. The main difference between the two is that, once built, we fix where the pipes will be transporting CO_2 to and from. Therefore we only decide how many pipes and what kind of pipes are built between clusters and sequestration sites, in addition to the capacity at which each pipe operates. We are assuming that all types of pipes can be built in all clusters. Thus we have

$$pd_t \in \mathbb{N}^{a\times(a+b)\times|S_p|} \times \mathbb{R}^{a\times(a+b)\times|S_p|}.$$

- $\mathbf{pd}_t^{(1)} \in \mathbb{N}^{a\times(a+b)\times|S_p|}$ indicates the number of pipes we build at stage t from a given cluster to another cluster or sequestration location. As with the ships we consider that the pipes available at stage t are those that have been built at stages $1,...,t-1$.
- The component $\mathbf{pd}_t^{(2)} \in \mathbb{R}^{a\times(a+b)\times|S_p|}$ indicates the capacity which is used during time stage t by pipes transporting CO_2 from a given cluster to another cluster or sequestration location. As for ships we consider that the change in time t affects the amount of CO_2 shipped in this step (and not the next one).

4.2.4 Sequestration decision

The final decision component is the storage decision. This decision presents an identical structure to the capture decision.

$$qd_t \in \mathbb{R}_+^b \times \mathbb{R}_+^b.$$

- Where $\mathbf{qd}_t^{(1)} \in \mathbb{R}_b^+$ tells us how many tonnes of CO_2 sequestration capacity we build at a storage site sq at time t. As with the rest of the infrastructure components, we assume this capacity becomes available the time step after it is built.
- Where $\mathbf{qd}_t^{(2)} \in \mathbb{R}_b^+$ tells us how many tonnes of CO_2 sequestration we choose to use at each storage site at time t. The amount we can use will be constrained by how much we have built is previous steps.

4.3 Stochastic Information

The stochastic data we consider is for $t \geq 1$:

$$\xi_t = C_t,$$

where $C_t \in \mathbb{R}_+^{a\times|K|}$ holds in its components the information regarding how many holds tonnes of CO_2 of each type are produced at step t for each cluster.

For the purposes of simulation, the following could be considered,

$$\xi_{t,cl} = \xi_{t-1,cl} + X_{t,cl}.$$

One could then choose to look at discretised versions of a variety of distributions, some examples of these could be:

1. $X_{t,cl}$ taking values from the set $\{-kC_0, 0, kC_0\}$ all with equal probability and where the family $\{X_{t,cl}\}$ are independent for all $t \in \{1,...,T\}, cl \in CL$.

2. $X_{t,cl} \sim N(0, \sigma t)$, where we take the parameter σt to be the variance in the CO_2 production from each cluster at time t and as before we consider the family $\{X_{t,cl}\}$ to be independent.

4.4 Objective function

As previously explained our model aims to maximize the amount of CO_2 stored at time T. Thus we have that the benefit obtained at time t is the total CO_2 stored at time t which is in turn

$$b_t \cdot x_t := \sum_{cl \in CL} \sum_{sq \in SQ} \left(\sum_{sp \in SP} pd^{(2)}_{t,cl,sq,sh} + \sum_{sh \in S_{sh}} sd^{(3)}_{t,cl,sq,sp} \right). \tag{1}$$

The left-hand term in the sum totals the amount of CO_2 taken from each capture location to each sequestration location through pipes. The right-hand term is analogous where now the CO_2 is transported through ships. Note that neither the stochastic information nor the previous decisions $(x_0, ..., x_{t-1})$ play a role in the definition of b_t. However, they will interact with our problem through the constraint set for x_t, and in particular the set where and $pd^{(2)}_{t,cl,sq,sh}, sd^{(3)}_{t,cl,sq,sp}$ must lie.

4.5 Constraints

We now discuss the constraints that govern the preceding variables. Encapsulated within these constraints are various assumptions we impose on our model. For this reason, we will carefully explain all the assumptions made and how they lead logically to each term in the equations that follow. In this way, we hope that the reader may gain full insight into the workings of the model, and thus, may be able to adjust them to any new data and information that may arise.

4.5.1 Capture, Transport and Sequestration Constraints

Firstly we present the constraints that determine the maximal theoretical capacity for CO_2 capture, transport, and storage. These are given by the following equations,

$$cd^{(2)}_{t,cl,k} \leq \sum_{s=0}^{t-1} cd^{(1)}_{s,cl,k} \quad \forall t, cl, k \tag{2}$$

$$cd^{(2)}_{t,cl,k} \leq C_{t,cl,k} \quad \forall t, cl, k \tag{3}$$

$$pd^{(2)}_{t,cl,cs,sp} \leq \sum_{s=0}^{t-1} pd^{(1)}_{s,cl,cs,sp} pc_{sp} \quad \forall t, cl, cs, sp \tag{4}$$

$$sd^{(3)}_{t,cl,cs,sh} \leq sd^{(2)}_{t,cl,cs,sp} sc_{sh} \quad \forall t, cl, cs, sp \tag{5}$$

$$qd^{(2)}_{t,sq} \leq \sum_{s=0}^{t-1} qd^{(1)}_{s,sq} \quad \forall t, cl, cs, sp \tag{6}$$

Equations (2), (3) express that we cannot capture more CO_2 than the capacity we have built-in previous steps and the amount of CO_2 a cluster produces allow respectively. Equations (4), (5) constrain the maximum amount of CO_2 that can be transported through pipes and ships respectively and (6) limits the amount of CO_2 which can be stored. Note that, due to the difference in the way the shipping decision is defined, the theoretical capacity for shipping is not expressed as a sum (this is different from what occurs with the theoretical bound for capture, pipeline transport, and sequestration). This occurs essentially because the mobility of ships makes it so that we decide where *all* ships are based in each step instead of just the ones we just built.

In interpreting the following constraints it is useful to visualize the model as a network formed of the pipes and ships built in prior time steps. This network is modified in each time step and determines how the CO_2 can flow within it and any associated costs that are incurred through operating and constructing it.

1. In our network, we cannot transport more CO_2 from an industrial cluster than the amount of CO_2 that it holds.

$$\sum_{k\in K} cd^{(2)}_{t,cl,k} + \sum_{cl'\in CL}\sum_{sp\in S_p} pd^{(2)}_{t,cl',cl,sp} + \sum_{cl'\in CL}\sum_{sh\in S_h} sd^{(3)}_{t,cl',cl,sh}$$
$$\geq \sum_{cs\in CL\cup SQ}\sum_{sp\in S_p} pd^{(2)}_{t,cl,cs,sp} + \sum_{cs\in CL\cup SQ}\sum_{sh\in S_h} sd^{(3)}_{t,cl,cs,sh} \quad \forall t, cl \quad (7)$$

The first summand indicates the quantity of CO_2 that is captured at the cluster cl, the second summand is the amount of CO_2 that is transported through pipes to cl through all other capture locations cl' and the third indicates the analogous situation for transport through ships. the fourth and fifth summands, on the right-hand side of the inequality, represent how much CO_2 is transported from cl through pipes and ships respectively to other clusters and sequestration sites.

2. The amount of CO_2 we store at each sequestration site is the sequestration capacity that is being used.

$$\sum_{cl\in CL}\left(\sum_{sp\in SP} pd^{(2)}_{t,cl,sq,sp} + \sum_{sh\in S_{sh}} sd^{(3)}_{t,cl,sq,sh}\right) = qd^{(2)}_{t,sq} \quad \forall t, sq \quad (8)$$

The sums on the left-hand side indicate the amount of CO_2 which is piped and shipped from all industrial clusters to a storage site sq. Whereas the right-hand side indicates the total capacity for storage which is usable at time step t (that is, which has been installed before stage t).

3. At stage t we can only use the ships that have been built at stages $1,\dots,t-1$.

$$\sum_{cs\in CL\cup SQ}\sum_{cl\in CL} sd^{(2)}_{t,cl,cs,sh} = \sum_{s=0}^{t-1} sd^{(1)}_{s,sh} \quad \forall sh \in S_{sh}, \quad \forall t \leq T \quad (9)$$

The first summand indicates the total number of ships of type sh that have been assigned for CO_2 transport and the second is the number of ships of this type that have been built in previous stages.

4. We have that the total amount of CO_2 sent to a sequestration site at the end of the duration of the project (left-hand side) cannot exceed the total capacity of the sequestration site.

$$\sum_{t=0}^{T}\sum_{cl\in CL}\left(\sum_{sp\in SP} pd^{(2)}_{s,cl,sq,sp} + \sum_{sh\in S_{sh}} sd^{(3)}_{t,cl,sq,sh}\right) \leq CA_{sq}, \quad \forall sq \quad (10)$$

The left-hand sum sums over t, the amount of CO_2 transported through pipes at time t to sq. Whilst the second summation is the analogous term for transport through ships. In practice we will set C_{sq} to be infinite as, since we are amalgamating various sequestration sites into a single one this capacity will never be surpassed. However we include this constraint in the case where a different approach is taken and sequestration sites are considered separately.

5. There can be no transport from an industrial cluster to itself.

$$pd^{(1)}_{t,cl,cl,sp} = pd^{(2)}_{t,cl,cl,sp} = sd^{(2)}_{s,cl,cl,sh} = sd^{(3)}_{t,cl,cl,sh} = 0, \quad \forall t, cl, sp, sh.$$

4.5.2 Budget Constraints

To be able to formulate our model within the framework of stochastic linear optimization we consider that all costs depend linearly on the decision variable. For example, the cost of building infrastructure for 10^6 tonne of CO_2 capture per year is ten times the cost of building 10^5 infrastructure for CO_2 storage. Therefore it is important to note that only considers economies of scale in an *averaged* sense. We say average because the prices of construction and operation that we take are from literature where authors already incorporate in some way economies of scale. Further work could investigate this problem through a formulation that could explicitly incorporate economies of scale. This could be done, for example, by discretizing some variables and associating to each discrete value a different cost which incorporates these economies of scale.

The budget constraints which we consider are therefore formulated as below.

1. The price of construction

$$PC_t = PCC_t + PCT_t + PQ_t.$$

Where PCC_t, PCT_t, PCQ_t are the price of construction for facilities for capture, transport and sequestration at stage t. We have that these terms are constructed as

$$PC_t = \sum_{k\in K}\sum_{cl\in CL} \lambda_{cc,k} cd_{t,cl,k} + \sum_{sh\in S_{sh}} scc_{sh} sd^{(1)}_{t,sh} + \sum_{sp\in S_p}\sum_{cl\in CL}\sum_{cs\in CL\cup SQ} pcc_{cl,cs,sp} pd^{(1)}_{t,cl,cs,sp} + qcc \sum_{sq\in SQ} qd^{(1)}_{t,sq} \quad (11)$$

2. The price of operating the facilities once built is,

$$PO_t = POC_t + POT_t + POS_t + POQ_t.$$

Where POC_t, POT_t, POQ_t are the operational price for infrastructure for the capture, transport, and sequestration at stage t. We take operating costs to be proportional to the amount of CO_2 captured and transported. In reality, some prices will be incurred even if the equipment is not being used, however, we consider this assumption to be reasonable and it is necessary in order to maintain linearity. Explicitly, we have

$$PO_t = \sum_{k\in K}\sum_{cl\in CL} \lambda_{oc,k} cd^{(2)}_{t,cl,k} + \sum_{sh\in S_{sh}}\sum_{cl\in CL}\sum_{cs\in CS} soc_{sh} sd^{(3)}_{t,cl,cs,sh} + \sum_{sp\in S_{sp}}\sum_{cl\in CL}\sum_{cs\in CS}\sum_{s=0}^{t} poc_{cl,cs,sp} pd^{(3)}_{s,cl,cs,sp} + qoc \sum_{sq\in SQ} qd^{(2)}_{t,sq} \quad (12)$$

Note that in particular, we assume that the operating cost of the ship is independent of the locations between which they travel. In practice, there may be some regional variations in labor costs but we count these as negligible to simplify our model. This said our price constraints are as follows:

$$0 \leq MC_t - PC_t \quad \forall t \tag{13}$$
$$0 \leq MO_t - PO_t \quad \forall t \tag{14}$$
$$0 \leq NO - PO_T \tag{15}$$

These constraints impose that the money used cannot surpass the budget supplied in each time step cannot. In particular we do not consider the case where money can be saved between time steps for future use. Though this could be another reasonable assumption. The first two equations (13)-(14) represent the need to stay within the CAPEX and OPEX budget for each time step whereas equation (15) represents the need to not have a final operational budget that is too high. We include this constraint so that the solution will not completely disregard operational costs after the prescribed final time T. The variable NO will be automatically assigned to take the value of the OPEX cost of the previous year.

4.6 Modelling Randomly Constrained CO_2 Capture

The above model presents the interest of determining optimal decisions in a framework in which capture, transportation, and sequestration are all considered simultaneously. However, in practice, it may be the case that the amount of CO_2 capture that can be built is limited by other factors. The model we propose above can be adopted without too much difficulty to cover the case where the cap on the amount of CO_2 we sequester comes from the limit on CO_2 capture.

To amend the model previously shown, one makes it such that the capture decision is no longer a variable. That is, our new decision at time t is

$$x_t = (sd_t, pd_t, qd_t).$$

The random information is now denoted by $\tilde{cd}_t$. This is instead of C_t (the CO_2 production) as in the the previous model. The random variable $\tilde{cd}_t$ will have its image in $\mathbb{R}_+^{a\cdot K}$ and $\tilde{cd}_{t,cl,k}$ will denote the amount of CO_2 capture capacity that is added at a cluster cl at time t from some industrial process k.

Equations (1)-(15) of the previous section now only need to be modified by:

- Swapping $cd^{(1)}_{t,cl,k}, cd^{(2)}_{t,cl,k}$ for $\tilde{cd}_{t,cl,k}$ in all the equations.
- Eliminating constraints (2), (3)

In our case, we will obtain the random variable $\tilde{cd}_t$ by using a range for conceivable CO_2 captures at each cluster. This data was obtained through expert solicitation and can be found in the github linked at the start of this report. The ranges in the data would lead to a scenario count of the order of 10^{30}. To properly incorporate this data we make the following simplifications.

- We suppose that, in each time step, the amount of CO_2 capture of each type k either increases by the difference between the maximum range of CO_2 capture in the previous time step and the maximum amount of CO_2 production of the next time ep. Alternatively, we suppose that it could stay the same as in the previous time step. If we write $m_{t,cl,k}$ for the maximum amount of CO_2 capture at time stage t at cluster cl of type k this would correspond to

$$\tilde{cd}_{t,cl,k} = X \cdot (m_{t,cl,k} - m_{t-1,cl,k}) \quad \forall t, cl, k.$$

Here $X \sim B(1/2)$ is a random variable following a Bernoulli distribution of parameter $1/2$ and we define $m_{-1,cl,k} := 0$.

- We link the CO_2 capture types of all the different clusters in the sense that we supposed that for each time step t either

$$\tilde{cd}_{t,cl,k} > 0 \quad \forall k, cl \quad or \quad \tilde{cd}_{t,cl,k} < 0 \quad \forall k, cl.$$

This would reflect the phenomenon that, if one kind of CO_2 capture increased in a certain cluster, then it would be somewhat reasonable to suppose that the reason for the increase could be extrapolated to all other clusters.

5 Model Specification

In order to be able to run our model, we must take values for all the parameters and sets that are provided in the previous part. Here we will explain where our data is coming from and how we have arrived at the assignment of different variables.

5.1 Clusters and Sequestration Sites

For this project, we will be taking the clusters and sites to be those identified in Figure 1,

$$CL = \{\text{St. Fergus, Grangemouth, Teeside, Humberside, Merseyside, South Wales}\}$$
$$SQ = \{\text{Nothern North Sea, Central North Sea, Southern North Sea, East Irish Sea}\}$$

We note again that we have simplified the setup of the sequestration sites as each section of the sea involves dozens of sites that each have unique capacities.

5.2 Shipping

In terms of ships, we base our prices off of the cost for low-pressure transport, which is the central case considered [2]. On page 25 of this article there is formula by which we can obtain the CAPEX of various different kinds of ships.

Based on the information on page 31 of this article we estimate that the ratio of OPEX for shipping (in our case fuel cost, harbor fees, and ship OPEX as we include liquefaction in the capture costs) to that of ship CAPEX is $34/14$ for a 20 year period. Assuming this ratio is fixed for each kind of ship we obtain from the CAPEX prices for each ship their respective OPEX prices.

To calculate the maximum amount of $\mathrm{CO_2}$ transfer of a given ship of type sp between two locations cl, cs we take

$$st_{cl,cs,sh} = \frac{\text{years} \times \text{operational hours per year}}{\left(2 \cdot \frac{d(cl,cs)}{\text{speed of ship}} + \text{unloading and loading time}\right)} \times \text{capship}$$

Where we take a time period of 5 years and the values of each of the other variables are taken from page 25 of [2] (8322h,15nm/h,55h) and the distance of the shipping routes between each port is taken from the appended excel sheet in [2]. We also need to calculate the distance of a shipping route between a port and a given sequestration site. To do so we use the coordinates of each port and the approximate central point of a sequestration site, then we will approximate the distance a ship would have to travel using Google Maps.

Using the maximum amount of $\mathrm{CO_2}$ a given ship can transport between two locations and knowing the OPEX per ship we calculate

$$soc_{cl,cs,sh} = \frac{\text{OPEX}_{sh}}{st_{cl,cs,sh}}$$

A final adjustment that is required is that of currency conversion. We have taken all these prices from £ in 2019 to £ in 2021 using Mathematica.

5.3 Piping

To obtain the price of different types of pipes, their capital cost per mile built and their capacity, we use the work of Heddle et. al [6] who estimate costs for onshore pipes based on the prices of natural gas pipes.

We take the OPEX cost to be OPEX = $\$3100$ yr/km, independently of pipeline diameter ([6] page 22). It is worth noting that prices do vary on this matter through the literature such as in Mikunda's work [9], where on page 2414 a price of £7000 yr/km is given. For the capacity, cap(sp), of each kind of pipe and their CAPEX prices per km, we use the information on [6] pages 19 and 22 respectively. Using this information we can derive the operating cost as,

$$poc_{cl,cs,sp} = \frac{\text{years} \times \text{OPEX} \times d(cl,cs)}{\text{cap}(sp)},$$

where the unit of the above term is $\$/MtCO_2$. Furthermore, it has been estimated that the price of offshore pipelines may be approximately 50% higher than onshore pipelines (see [9] page 2414). Thus we multiply the CAPEX price by a factor of 1.5 as we only consider offshore pipelines. Finally, these prices are in \$ for 2003, so we transfer them to £ for 2021.

5.4 Capture and Sequestration Costs

As for the price of capture we set the CAPEX for direct air capture to be $\$780$ in 2018 dollars for a plant with a capture capacity of $1MtCO_2$/year based on Keith's calculations [8] (page 1589). To calculate the operation costs related to capturing $1tCO_2$/year we take the mean of the estimate of $\$97-232$ year/tCO_2 and convert these amounts from \$ in 2018 to £ in 2021.

For the medium cost CO_2 we use a price of capture for $1tCO_2$ to be equal to $\$74$ in 2013 based on the cost analysis for NGCC powerplants found in page 5 of Rubin's work [11]. We assign the cheap capture option's OPEX to be half of this per tonne of CO_2. Based on the results on [10] (page 27) we set a price of £30/MWh for the CAPEX costs of capture for both the medium and low-cost CO_2 capture. In this same article, a cost of £$37/tCO_2$ captured and of £75/MwH is estimated. Thus we use as a rough approximation a cost of $2tC02/Mwh$ captured which leads us to an estimate of £$150/tCO_2$ captured for CAPEX.

To all of these capture prices, we add a conditioning cost of 5.3 €$/tCO_2$ based on [10] page 6 (these are liquefaction costs but for simplicity, we estimate a similar cost for conditioning of CO_2 for pipeline transport).

As for sequestration costs we use the CAPEX of $120M$ € and OPEX of $6M$ €/year based on the prices given for an offshore DOGF with $66MTCO_2$ capacity in [10] page 25. From here we set

$$\text{qcc} = \frac{120}{66} M€/MtCO_2; \quad \text{qoc} = \frac{6 * years}{66} M€/MtCO_2.$$

Both the sequestration and conditioning prices are in 2011 € and we convert them to £ in 2021 using Mathematica.

Part IV
Simulations and Results

"'Obvious' is the most dangerous word in mathematics."

Eric Temple Bell

Now that we have constructed a model to look to build the optimal infrastructure, we will utilise this part of the report to explore how we then built a program to simulate and obtain solutions to this model. Initially, when we built the model we were looking at the case where CO_2 production was stochastic. Upon receiving the data for our model we changed this to look at the case where instead it is CO_2 capture was instead stochastic. We also introduce a nodal formulation of the model which allows us to recover optimal decisions for $t > 0$. Our implementation in the programme FICO Xpress led to many large scenario cases. Therefore we will also explore the techniques and thought processes that were implemented for the programme to work quickly and interestingly.

6 Xpress and Computational Complexity

For this project, to write the programme, we were advised to use FICO Xpress. This is a software package where the user writes code in a language called Mosel and an optimization can be ran. After initially writing a deterministic version of the model into Xpress, which we were able to run well, we looked to implement the stochastic element into the model. Unfortunately, we found that Xpress no longer had a maintained package that allowed variables to be declared as stochastic and this led us to have to rethink how we would implement our programme efficiently. An extended deterministic model which in some way incorporated probabilities was the necessary implementation to make, however led to the question of how to stop the model from anticipating the perfect future of the model. The work of Ahmed [1] informed our choice to use non-anticipativity constraints and use a nodal formulation within the model build.

6.1 Nodal Formulation

As was noted previously, the Multistage Stochastic Linear Programming (MSSLP) only returns the optimal initial decision x_0 and subsequent decisions are ill-defined. To circumvent this issue we take advantage of the fact that our random variables are finitely distributed. The crucial idea is to view solutions as functions of the possible scenarios. Since our random variables are discrete we have that for each t the random information ξ_t takes values in the finite set A_t. Let us set

$$\mathcal{T} := A_1 \times \cdots \times A_T, \quad x = (x_1, ..., x_T), \quad \xi := (\xi_1, ..., \xi_T),$$

and consider the constraint set

$$S(x, \xi) := S_1(x_0, \xi_1) \times \cdots \times S_T(x_0, ..., x_{t-1}, \xi_T) \subset \mathbb{R}^{d \times T}.$$

We also define the set of *scenarios equal up to time* t to be the set of pairs

$$\mathcal{I}_t := \{(sc, sc') \in \mathcal{T}^2 | sc_i = sc'_i, i = \{0, ..., t\}\}.$$

Where the constraint set is determined by the same constraints as in the previous sections. Then we consider the *nodal formulation*

$$\arg\max_{x} \sum_{sc \in \mathscr{T}} \left(\mathbb{P}(\xi = sc) \cdot \sum_{t=0}^{T} b_t x_t^{sc} \right) \tag{16}$$

$$x \in S(x, sc) \quad t = 0, \ldots, T \tag{17}$$

$$x_t^{sc} = x_t^{sc'} \quad \forall (sc, sc') \in \mathscr{I}_t. \tag{18}$$

Where it is important to note that, the object x that is a solution to our problem,

$$x : \mathscr{T} \to \mathbb{R}^{d \times T}; \quad x(sc) \to (x_0^{sc}, \ldots, x_T^{sc}),$$

is now a function of the scenarios. Additionally we comment on the appearance of the extra *nonanticipativity* constraint in (18). This constraint stipulates that decisions cannot be made as if we knew the future, that is, we must act equally up to time t if we observe the same reality up to time t.

Equations (16)-(18) tell us for each scenario the optimal decisions not only at step 0 but also all the way to time step T and are thus the equations we wish to solve. Additionally, we comment on the appearance of the extra *non-anticipativity* constraint in (18). This constraint stipulates that decisions cannot be made as if we knew the future, that is, we must act equally up to time t if we observe the same reality up to time t. Note that the non-anticipativity constraint (18) may add considerable computational complexity to out problem and in the next sections we describe how to efficiently implement this constraint.

7 Stochastic Programme Build

In our initial simulations, though we had eliminated the need to include the CAPEX for carbon capture we still included the OPEX for capture. In this situation we saw that our model with the nodal formulation ran very quickly. The output of the model however was not very interesting as essentially all the money was bein spent to be able to pay for the CO_2 capture and not to construct the network of transportation.

After some consideration, we removed the OPEX cost as well and this led to the programme running for a much longer time. This was due to it now being able to make more complex decisions with regards to transportation without the budget being engulfed in capture OPEX costs. We'd now gone to the opposite end of the spectrum in terms of run time and therefore needed to look into ways in which we could reduce the number of scenarios that the programme was considering.

7.1 Reduction of Computational Complexity

In order to reduce the complexity, we considered two aspects of the programme: the number of decision variables and the number of constraints.

Firstly, we reduced the number of decision variables. To do this we used what is known in Mosel as dynamic arrays. Using this datatype it is possible to only define the variables that you actually need. In our case, this means that we did not define the decision variables for piping and building pipes in the locations where our data says that it is not possible.

Secondly, we reduced the number of constraints. To do this we focused on the non-anticipativity constraints, as these were the majority of the constraints. Initially, we implemented these constraints as follows: for each time t we check all pairs of scenarios. If the scenarios are equal up to time t, we constrain the decisions at that time to be equal. This results in a collection of dense networks of constraints, as visualised in the left hand side of Figure 2. To make this more efficient, we did the following; for each time t we divide the scenarios into sets that are equal up to time t and choose a representative for each of them. To do this we use a method where we do not need to compare each pair, but use properties of the representation of the scenarios that we use in the Mosel code. We then give constraints in the form where we set the decisions in time t equal to the decision of the representative scenario of the collection the scenario is in. This results in star networks of constraints as visualised on the right-hand side of Figure 2.

7.2 CAPEX vs OPEX

Initially, we had set the OPEX to CAPEX ratio within the model to 5%, as an IEA report [7] indicated that this was the percentage of OPEX that could be attributed to annual CAPEX. However, running the model with this assumption it became clear that the CO_2 capture was far from optimal. This led to us simulating the model with the ratio of OPEX to CAPEX variable for a fixed budget. The results of this simulation showed that a ratio of 40% CAPEX and 60% OPEX were optimal. For this reason, in future simulations, we used these values.

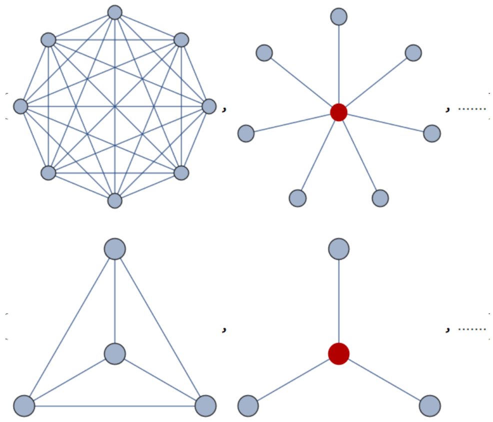


Figure 2: Constraint Networks

7.3 Visulation of Results

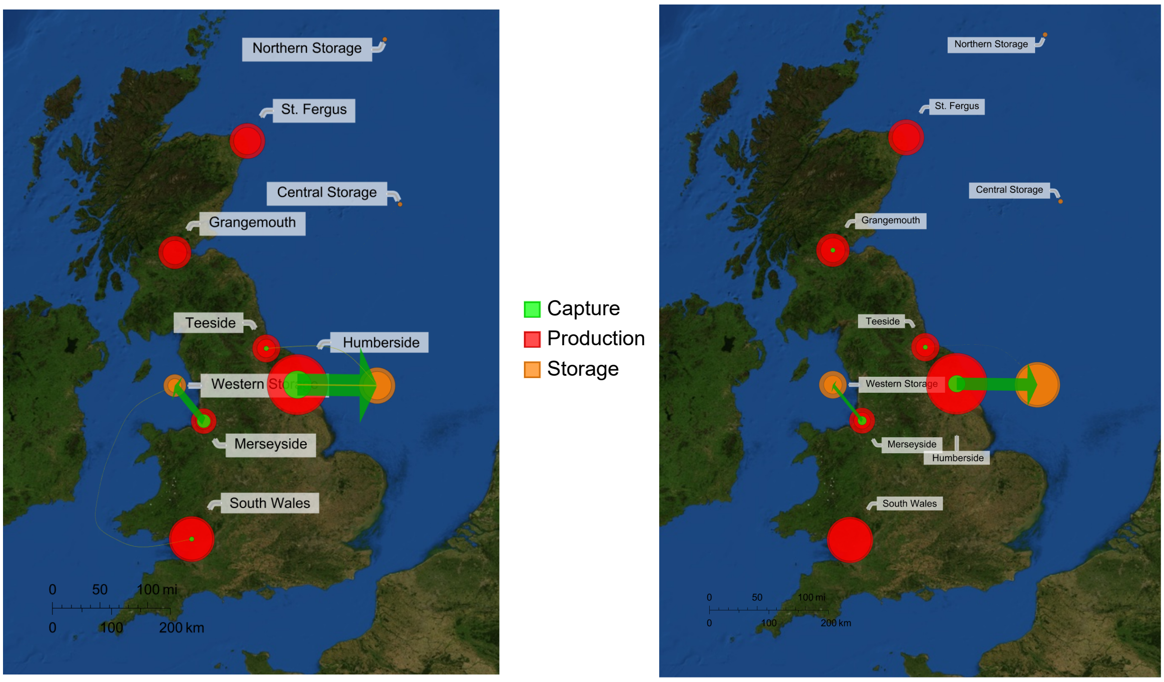
To complete this section, we now present a visualisation of the results run for three-time steps where we set an OPEX and CAPEX of

$$MC = (200, 100, 0); \quad MO = (0, 100, 350)$$

in million £ for each time step. The time span of the model corresponds to $2025-2050$ were the first two time steps correspond to five year periods and the last one to 15 years. The reason why an OPEX budget is incrementally added towards the end of the simulation is that no OPEX need to be used in the initial time step where all the infrastructure is still being built.

Below we present figure 3 showing the expected capture, production and sequestration rate of CO_2 per year in the final two-time steps (in the initial time step all the infrastructure is still being set up and thus, no CO_2 can stored). The green arrows correspond to CO_2 transported through pipes and the (very faint) yellow to CO_2 transported through ships. As we can see the model presents a strong preference for pipe transport and transports CO_2 between nearby clusters to sequestration sites. We note further that no transport between two different industrial clusters is used despite the model allowing for it.

We also include below a bar chart, figure 4, showing the expected amount of CO_2 sequestered from each cluster during each of the last two time steps, and where the first bar corresponds to the total amount of CO_2 available for sequestration. We note that with the budget supplied the only 13.6% of the CO_2 captured is sequestered. This suggests the need for a larger budget to be able to capture a satisfactory amount of CO_2 .



(a) $2030-2035$ (b) $2035-2050$

Figure 3: Rates of CO_2 capture, transportation and storage $t = 1, 2$

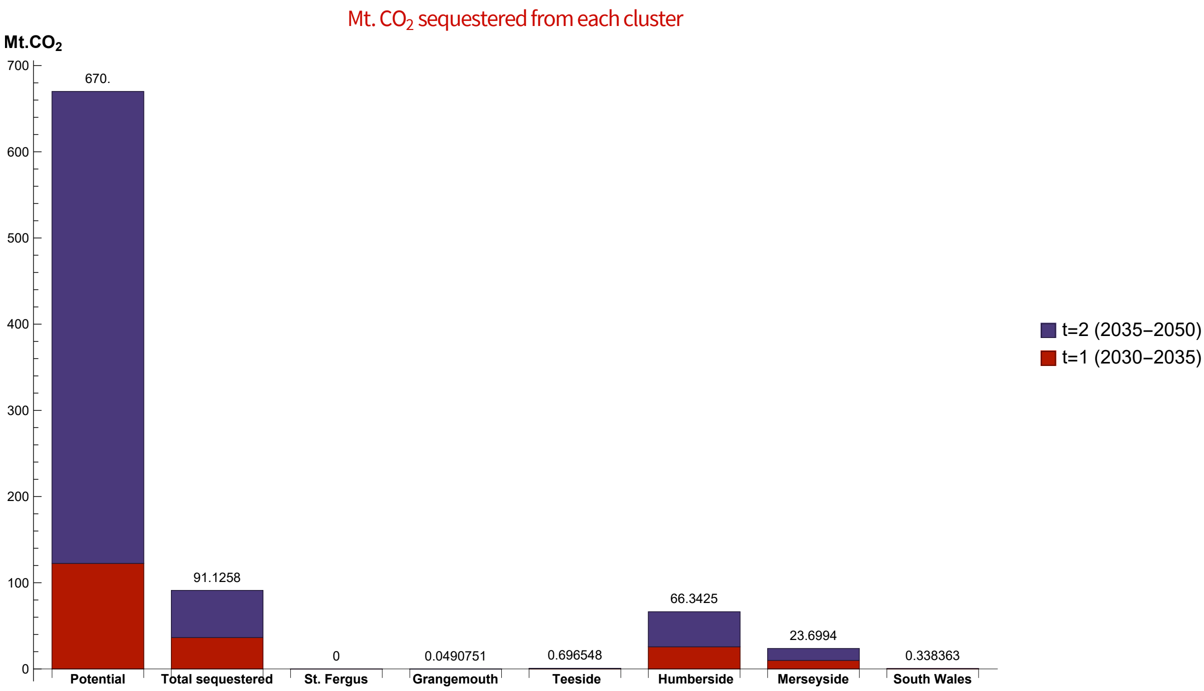


Figure 4: Bar Chart Map Capture

8 Sensitivity Analysis

In the following section, we introduce some measures that may be used to analyze the stability of the model under small changes in some of the parameters. Though we did not have time to implement these tools in their work they could be an interesting part of a future analysis.

We first introduce two parameters λ, r which will serve to parametrize:

- Different sizes of the total budget which we parametrize by $\lambda \in [1-\epsilon, 1+\epsilon]$.
- Different ratios of OPEX to CAPEX which we parametrize by $r \in \left[\frac{1}{2}-\epsilon, \frac{1}{2}+\epsilon\right]$.

Where the parametrization is given by

$$MC_t(\lambda) := \lambda MC_t(1); \quad MO_t(\lambda) := \lambda MO_t(1); \quad \frac{MO_t(\lambda, r)}{MC_t(\lambda) + MO_t(\lambda, r)} := r.$$

That is

$$MO_t(\lambda, r) = \frac{r MC_t(\lambda)}{(1-r)}$$

Consider a decision $x(\lambda, r, \xi)$ for some values of λ, r and note that this decision is a random variable as it depends on the realization of ξ. Then we wish to study the behaviour of

1. The *decision metric*

$$||x(\lambda, r, \xi)||_\infty := \sum_{t=0}^{T} \left(\sum_{cl \in CL} \sum_{cs \in CL \cup SQ} \sum_{sh \in S_h} \left| sd^{(2)}_{t,cl,cs,sh} \right| + \sum_{cl \in CL} \sum_{cs \in CL \cup SQ} \sum_{sp \in S_p} \left| pd^{(1)}_{t,cl,cs,sh} \right| + \sum_{sq \in SQ} \left| qd^{(1)}_{t,sq} \right| \right).$$

   Where for simplicity in the expression we omit the dependency on λ, r, ξ of the right hand side of the above equation. This metric aims to measure the total contribution of the decisions overall time that are made. Note that we only include in this norm the component of each decision that measures how we allocate resources. This is because these terms all have the same units ($t CO_2$) and represent effectively how we build infrastructure. It could be interested to calculate

$$m(\lambda, r) := \mathbb{E}_\xi \left[||x(\lambda, r, \xi) - x(\lambda_0, r_0, \xi)||_\infty \right].$$

   Where x_0 is taken as a base case in which the CO_2 production remains unchanged throughout time over all clusters. Note that the integral in the cases we simulate reduces to a finite sum weighted by the probabilities of the different scenarios.

2. The average amount of CO_2 sequestered over each sequestration location and the total amount of CO_2 sequestered

$$coq_{sq}(\lambda, r) := \sum_{t=0}^{T} \mathbb{E}_\xi \left[qd^{(2)}_{t,sq}(\lambda, r, \xi) \right]; \quad COQ(\lambda, r) := \sum_{t=0}^{T} \sum_{sq \in SQ} \mathbb{E}_\xi \left[qd^{(2)}_{t,sq}(\lambda, r, \xi) \right] \tag{19}$$

3. The average amount of CO_2 captured at each industrial cluster and the average amount of each type captured (easy, medium, direct air capture)

$$coc_{cl}(\lambda, r) := \sum_{t=0}^{T} \sum_{k \in K} \mathbb{E}_\xi \left[cd^{(2)}_{t,cl,k}(\lambda, r, \xi) \right] \tag{20}$$

$$coc_k(\lambda, r) := \sum_{t=0}^{T} \sum_{cl \in CL} \mathbb{E}_\xi \left[cd^{(2)}_{t,cl,k}(\lambda, r, \xi) \right] \tag{21}$$

4. The average amount of CO_2 piped and shipped

$$cos(\lambda, r) := \sum_{t=0}^{T} \sum_{cl \in CL} \sum_{cs \in CL \cup SQ} \sum_{sh \in S_{sh}} \mathbb{E}_\xi[sd^{(3)}_{t,cl,cs,sh}(\lambda, r, \xi_\mu)] \tag{22}$$

$$cop(\lambda, r) := \sum_{t=0}^{T} \sum_{cl \in CL} \sum_{cs \in CL \cup SQ} \sum_{sp \in SP} \mathbb{E}_\xi[pd^{(2)}_{t,cl,cs,sp}(\lambda, r, \xi)] \tag{23}$$

5. For some *fixed* representative values of λ, r it would be interesting to show as was done in Figure 3 a visualization of the average quantity of CO_2 captured, transported through pipes and ships, and sequestered throughout time. That is,

$$coc_{t,cl}(\lambda, r) := \sum_{k \in K} \mathbb{E}_\xi\left[cd^{(2)}_{t,cl,k}(\lambda, r, \xi)\right] \tag{24}$$

$$cos_{t,cl,cs}(\lambda, r) := \sum_{sh \in S_{sh}} \mathbb{E}_\xi[sd^{(3)}_{t,cl,cs,sh}(\lambda, r, \xi)] \quad \forall t, cl, cs \tag{25}$$

$$cop_{t,cl,cs}(\lambda, r) := \sum_{sp \in SP} \mathbb{E}_\xi[pd^{(2)}_{t,cl,cs,sp}(\lambda, r, \xi)] \quad \forall t, cl, cs \tag{26}$$

$$coq_{t,sq}(\lambda, r) := \mathbb{E}_\xi\left[qd^{(2)}_{t,sq}(\lambda, r, \xi)\right] \tag{27}$$

Part V
Conclusions and Discussion

"The essence of mathematics lies in its freedom."

Georg Cantor

The climate crisis leaves humanity no choice but to implement decisions to decrease our carbon footprint. CCUS has been identified as an important method of doing so and motivates the need for studies to determine its optimal implementation. In this work, we have considered a model that determines this optimal implementation while allowing for the presence of unknown information in the form of random variables for CO_2 emissions and, separately, for CO_2 capture.

Simulations of the model were run in this second case and showed that a larger OPEX than CAPEX will be needed than what was found in the literatur to be able to support an effective roll-out. This consideration is important when it comes to policy decisions as to budget allocation and must be taken into account. Furthermore, it was found that for a total budget of $£750M$ only a small percentage of CO_2 was captured, even without accounting for the price of CO_2 capture and storage which will be substantially higher than the cost of transportation and sequestration. This suggests that a larger budget than the $£1.3$ billion proposed by the government [12] is required. Future work could also look to determine this number, the stochastic model we propose allows for this by simply incorporating the budget as a decision variable.

It is important to note that throughout this report we have constructed a model which has made many assumptions with regards to prices, infrastructure, and political decisions. These assumptions have been detailed in full with the hopes that a future study could build upon them. The points below form the closing of what only looks to be the beginning of how we can use mathematics to optimize infrastructure for CO_2 removal and how future work may look to refine the modelling and simulations undertaken within this report.

Simplification Unpacking

For future work, a starting point could be to consider the clusters and sequestration sites in a more fine-grained fashion. In our work, we've made the simplification of considering 6 clusters and 4 sequestration sites. Each of these can be split into several production sites and sequestration regions. Furthermore, for reasons of computational complexity, it was necessary to link CO_2 capture from different industrial clusters. It would be interesting to try to implement a model in which some of these simplifications are unnecessary and compare any obtained results with the model proposed here. Further, in terms of sequestration, we have made the assumption that there will always be space somewhere in the region. However, at individual sites within each part of the sea, it is often not possible to determine the exact capacity of a site until the CO_2 starts to be sequestered. Thus, it could be interesting to introduce another random variable to control the capacity available at each of these sequestration sites.

Another key simplification that could be unpacked, would be that of the time periods. Here we have modelled 5 year time gaps between the moment infrastructure begins being built and the time its construction is finalised. The details of what is happening during that time are not elucidated. In reality, however different infrastructure could take a variety of times to build and it may be expedient to introduce a time step of a different size to account for this. Finally economies of scale were only considered in an averaged sense and an implementation that looked to include these explicitly could be considered in the future.

Political-Mathematical Complexities

As a mathematician, it is often our role to become as objective as possible when building a model. Within this report we have allowed more recent political decisions, namely that of the decision of where CCUS will be deployed first, to be filtered into the simulations and data. The stability of the model and how the infrastructure is optimized for deployment will always depend on the budgeting constraints. This amount in the real world could become very varied based on the environmental interests of whichever political parties are in power within the country at the time. In addition to this, the idea of Scottish independence could loom and potentially take place before the end of implementation in 2050. This could raise questions about which clusters can still be linked and will have funding.

We hope that within this report objectivity has been conserved and could withstand the changes that may happen as described above. It is not without note however that this model has been built with the intention that CCUS will be deployed with full backing and intentions to follow through. This perhaps may have caused minor bias in terms of implementation should a decision to volte-farce on usage of CCUS take place.

One further consideration we believe would be important to extend the model, would be to look at what a reasonable budget would be in order for CCUS capture costs to be included. When we initially did run our model with the CAPEX and OPEX costs associated with the capture of $\mathrm{CO_2}$ included, we saw that not much $\mathrm{CO_2}$ would end up being sequestered. This is because the building and maintenance of the initial capture costs would be so high that the budget would be engulfed. This means that one could potentially 'rejig' the constraints that we have set within the model and the programme to instead have a minimum threshold of $\mathrm{CO_2}$ that must be captured and sequestered and see what budget would be required for this.

The Human Touch

One final item that we have considered along the journey of this project and within this report, something which we believe is vital to the success of optimal infrastructure planning, is that there must always be some sort of human touch when modelling. Throughout there have been moments when intuition has kicked in, such as looking to change CAPEX to OPEX ratios, where we have been able to make improvements to the model by the information that has been fed in. Throughout future planning of CCUS implementation, human oversight must remain part of the process, with the model being further and further defined. This will help with interpretation of results and allow for accountability of the model usage. Why is this important? Quite simply because if left to run without interpretation and refinement, some poor decisions could be made around the infrastructure build of CCUS. All in the name of a net-zero carbon future.

References

[1] AHMED, S. *Stochastic Optimization*, chap. 29, pp. 379–391. Mathematical Optimization Society and the Society for Industrial and Applied Mathematics, 1st edn. (2017). doi: 10.1137/1.9781611974683.ch29.

[2] BEISD. Shipping $c0_2$-uk cost estimation study (2018). [Online; accessed 7-October-2021]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/761762/BEIS_Shipping_CO2.pdf.

[3] COMMITTEE ON CLIMATE CHANGE. Reducing uk emissions– 2018 progress report to parliament. (2018). [Online; accessed 7-October-2021]. Available from: https://www.theccc.org.uk/publication/reducing-uk-emissions-2018-progress-report-to-parliament/.

[4] DANTZIG, G. *Linear programming and extensions*. Rand Corporation Research Study. Princeton Univ. Press (1963). doi:10.7249/R366.

[5] DEPARTMENT FOR BUSINESS, ENERGY & INDUSTRIAL STRATEGY. The uk carbon capture, usage and storage (ccus) deployment pathway: an action plan (2018). [Online; accessed 7-October-2021]. Available from: https://www.gov.uk/government/publications/the-uk-carbon-capture-usage-and-storage-ccus-deployment-pathway-an-action-plan.

[6] HEDDLE, G., HERZOG, H., AND KLETT, M. The economics of co2 storage. *Massachusetts Institute of Technology, Laboratory for Energy and the Environment*, (2003).

[7] IEA. The future of hydrogen. (2019). Accessed: 2021-12-01. Available from: https://www.iea.org/reports/the-future-of-hydrogen.

[8] KEITH, D. W., HOLMES, G., ANGELO, D. S., AND HEIDEL, K. A process for capturing co2 from the atmosphere. *Joule*, **2** (2018), 1573.

[9] MIKUNDA, T., VAN DEURZEN, J., SEEBREGTS, A., KERSSEMAKERS, K., TETTEROO, M., AND BUIT, L. Towards a co2 infrastructure in north-western europe: Legalities, costs and organizational aspects. *Energy Procedia*, **4** (2011), 2409.

[10] PLATFORM, Z. E. The costs of co2 capture, transport and storage. *Post-demonstration CCS in the EU. European Technology Platform for Zero Emission Fossil Fuel Power Plants*, (2011).

[11] RUBIN, E. S., DAVISON, J. E., AND HERZOG, H. J. The cost of co2 capture and storage. *International Journal of Greenhouse gas control*, **40** (2015), 378.

[12] TREASURY, H. Autumn budget and spending review 2021. (2021). Accessed: 2021-12-13. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1029973/Budget_AB2021_Print.pdf.