**CCS: Optimization of Source-Sink Network Under Uncertainty**

**Introduction**

Off-shore carbon capture and sequestration (CCS) has not undergone a full cost assessment, considering the uncertainties and suitability of different capture, transportation and storage options. The project aims to create an ArcGIS model for generation of potential source-sink networks, to aid in the evaluation of uncertainties in early stage CCS projects.

The Intergovernmental Panel on Climate Change (IPCC, 2014) describes the need for CO Capture and Storage as part of the many possible global mitigation models if the 2°C target is to be met. Without this technology, the IPCC estimates the costs of reaching this goal could be 138% higher. However, the success of CCS, at scale, depends on a number of factors including financial incentives, assessment of storage risk, and public perception. The research proposed here seeks to address some of the multi-faceted challenges for this technology.

Currently, many CCS projects exist around the world with both pilot projects and commercial scale operations, which capture CO$_2$ from natural gas production, at Sleipner, Snøhvit and In Salah (Michael et al., 2010). These projects have helped to test injection as well as monitoring and verification technologies. Carbon dioxide has also been used for many years in tertiary, or enhanced, oil recovery, with the first long distance CO$_2$ pipeline built in the 1970s carrying 40 MtCO$_2$ per year to sites mainly in Texas (IPCC, 2005). However, for CCS to play an important role in future reductions of CO$_2$ emissions, further development and optimization of capture from anthropogenic sources and transport to sites for geologic sequestration need to occur.

**Methodology and Analysis**

Previous work has been conducted by Middleton and Bielicki (2009), which used GIS and mixed-integer linear programming to determine an optimal network of sources to sinks. However, the analysis presented was atemporal, with all infrastructure being built simultaneously. By considering a two stage problem, the
analysis is split up into two time periods. The optimal solution for time period 1 is optimized given the uncertainties in the variables for time period 2. For example, should the cheapest source-sink network be built now? Or should a costlier network, with redundancy and spare capacity, be built given there may be demand for more CCS in the future. The analysis requires information on the costs associated with the sources of CO₂, sinks (reservoirs for storage) and transportation network.

**Sinks**

The Gulf Coast Carbon Center, at the University of Texas’ Bureau of Economic Geology, are currently working on analyzing and modeling 8000 square miles of prospective geologic storage resources in Texas State waters. The abundance of data from the extensive characterization of the area by the oil and gas industry lends itself to suitability assessments for CO₂ storage, with data such as well logs, cores and seismic being used (Meckel et. al 2017). The analysis of the data aims to identify sites with the potential to store over 30 million metric tons of CO₂.

**Sources**

This region of the Texas Gulf Coast is also known to contain many point sources of carbon dioxide. In particular, these include sources that already implement capture technology and have a high purity CO₂ stream as a result. This reduces the capital costs for implementing CCS as the off-gas can be dehydrated, compressed and then transported for storage (de Coninck et al. 2010). Such sources include production of hydrogen, ethanol, ethylene oxide, ammonia, along with natural gas and LNG processing and gasification.

To identify candidate sources of CO₂ the EPA’s Greenhouse Gas Reporting Program (GHGRP) was compared to the data for point sources in the National Carbon Sequestration Database and Geographic Information System (NATCARB). The EPA’s GHGRP requires large greenhouse gas (GHG) emission sources to be reported. These requirements apply to source emissions exceeding 25,000 metric tons CO₂e per year; supply of products which if released, combusted or oxidized would release over 25,000 CO₂e per year; or facilities which received over 25,000 metric tons CO₂ per year for underground injection (40
CFR part 98). The total emissions reported are attributed to a particular category depending on the process from which the CO₂e originated. In total there are 41 separate categories which can be listed as the process origin of the emissions.


NATCARB data was downloaded (National Energy Technology Laboratory, 2015) and loaded into ArcMap 10.3.1. The point sources between Freeport, TX and Lake Charles, LA, were selected and downloaded. NATCARB data was used as the template for compiling a source database to ensure continuity between the current study and the database previously compiled by the U.S. Department of Energy. The most recent data, 2015, from the EPA’s GHGRP was downloaded for the applicable counties between Freeport, TX and Lake Charles, LA using the Facility Level Information on Greenhouse gases Tool, FLIGHT (United States Environmental Protection Agency, 2016).

A series of macros were developed to compare the NATCARB and EPA GHGRP data using the unique facility identifier (GHGRP_ID) and create an updated point source database for the northeastern Gulf Coast. The emissions were updated to those from 2015 and the historical emissions dating back to 2010 were scraped from the EPA Facility Level Detailed Reports (United States Environmental Protection Agency, 2016) and appended to the spreadsheet. In addition, information relating to percentage ownership and zip code of the facilities was obtained from the reports using a separate macro. Finally, data relating to the emissions for each applicable subpart were scraped from the Detailed Facility Report for each entry in the spreadsheet. This enabled the volume of high purity off-gas from the facilities to be determined. Of particular interest were emissions from Subpart G - Ammonia Manufacturing, Subpart J – Ethanol Production, Subpart P - Hydrogen Production, Subpart W - Petroleum and Natural Gas Systems and Subpart X - Petrochemical Production. For broad categories, such as Subpart X, the macro also searched for words
such as ‘ethylene oxide’ or ‘acid gas removal’ to find any specific emissions within the subpart that could be high purity streams.

Once the data in NATCARB had been updated and aligned with the current GHGRP reporting year, other data sources were investigated. The ‘Emissions by plant from CO₂, SO₂, and NOx’ (United States Energy Information Administration, 2015), ‘Emissions and Generation Resource Integrated Database (eGRID)’ (United States Environmental Protection Agency, 2017a) and ‘EIA-860 Detailed Data’ (United States Energy Information Administration, 2017) were compared to the database. These datasets pertain to emissions from electricity generators specifically. However, this data is still useful in compiling a database of CO₂ point sources. The report ‘EIA-860 Detailed Data’ also shed light on which of these facilities used acid gas removal on the outlet gas streams.

Once the database had been compiled, the latitude and longitude for each source in the area of interest was verified using a combination of Google Earth and Google Maps. Aerial imagery for the facilities with over 100,000 metric tons total CO₂e per year was saved and the availability of land for retro-fitting capture technology at the site was recorded.

Searches of the TCEQ (Texas Commission for Environmental Quality) and FERC (Federal Energy Regulatory Commission) were then used to determine the capture technologies at a selection of the high-purity sources identified during the compilation of the database. The different technologies identified were:

- Rectisol – physical solvent
- Methanolamine – chemical solvent
- Methyl-diethanolamine – chemical solvent
- Potassium carbonate – chemical solvent
- Pressure swing adsorption
The method of CO₂ removal was important to determine as each has different operating, and regeneration, temperatures and pressures as well as capture efficiencies, which impacts the overall system cost (National Energy Technology Laboratory). The analysis of permits and applications also highlighted which facilities directly emitted this off-gas and those who recycled or sold the carbon dioxide for other processes. Sources currently in the application and pre-filing stages were also identified during this step. Detailed calculations and process flow diagrams are made available before any permits to emit are issued, allowing assumptions on potential emissions to be made. The selected high purity sources in Houston and Port Arthur are shown in Figure 1 and Figure 2.

**Transportation**

Once the sources and sinks were identified, a standalone model was created in ArcMap to provide a network of least cost routes from each source to each sink. Three sources and two sinks were used to demonstrate the model. However, it can be run with any number of each.

The model provides a tool for the generation of a network of CO₂ pipelines for anywhere within the USA. It consists of nodes (including the sources and sinks) along with arcs. Each arc has a cost associated with it, allowing for the network to be optimized as a modified transportation problem. The cost weights for CO₂ transportation via pipeline are defined in the paper by MIT (2006). These costs are dependent on the following information, contained in the geodatabase for the contiguous United States, for:

- NHDPlus Flowlines and Waterbodies (National Hydrography Dataset, 2017)
- Railroads (United States Department of Transportation, 2017)
- Roads (United States Census Bureau Department of Commerce, 2015)
- Elevation (ArcGIS Elevation Server)
- National Parks (United States Department of the Interior, 2017)

The model requires the user to input the following information:
Land Cover Raster Data for the State of interest (United States Geological Survey, 2011)
  o Placed into the NLCD folder

State Park Feature Class for the State of interest (Texas Parks and Wildlife Division, 2015)
  o Placed into the Geodatabase

Source locations
  o CSV file

Sink Locations Point Feature Class

Once the required data is loaded into the geodatabase, the user can run the tool. They are prompted to enter the cell size, in meters, and select the areal extent of the analysis. Larger cell sizes are suitable for larger analysis areas in order to reduce the computation time for the model. The inputs can be seen in Figure 3.

The model was created in a hierarchical structure, Figure 4, which includes a top-level model and 10 sub-models. This structure was chosen as ModelBuilder only allows one iterator per model. Additionally, such a structure enables easier modification and debugging.

**Top Level – Level 0**

The top level model aggregates the sub-level models and orders them. It also creates three parameters which the user will enter: the coordinate system, cell size and extent. For the extent, a blank feature class was placed into the geodatabase. In the model, this feature class was changed to a data type ‘Feature Set’. This allows the user to select their area of interest.

The top level model also implements ‘Create Feature Dataset’. The feature dataset is created so that when the feature classes are copied into it their coordinate systems are unified. This allows alignment of the cells when they are subsequently converted to rasters and calculations are performed. See Figure 7.

**Extract Rasters – Level 1**
The ‘Extract Rasters Model’ is shown in Figure 8. The two rasters, the National Land Cover and Elevation Datasets, are extracted to the extent defined by the user. This reduces the size of the raster to be used in the computations from the national scale. The National Land Cover Dataset is reclassified as shown in Table 1.

The slope is calculated from the National Elevation dataset and reclassified, Table 2. In both cases, the original extracted raster is then deleted to reduce unnecessary files in the geodatabase. Once this is completed the sub-model is run.

**Resample – Level 2**

The Resample model, Figure 9, iterates through the reclassified NLCD and slope rasters created in the parent model and resamples them to ensure they have the same cell size as the user specified when running the model. This simplifies the raster calculations by ensuring cells overlay each other directly. The resampled rasters for all the iterations are collected into a single location.

**Clip Features – Level 1**

The model iterates through all the feature classes in the geodatabase and clips them to the user input area of interest. The clipped feature classes are then copied to the feature dataset ‘Base’ created in the ‘Top Level’ model. This ensures they all have the same coordinate system. Once the feature classes have been copied, the clipped feature classes are deleted to remove unnecessary files from the geodatabase.

Two ‘If-Else’ statements were implemented in the model using the Calculate Value tool. These statements utilize expressions in Python to return true or false as a precondition to subsequent steps. One determines if the feature class operated on during the current iteration loop is ‘Clip Poly’. If it is, the copied feature class is also deleted as ‘Clip Poly’ isn’t required in subsequent steps. The second ‘If-Else’ statement determines if the feature class operated on during the current iteration loop is ‘NHDPlus Flowlines’. If it is then the ‘Select’ tool is used to select all features that don’t have type
‘pipeline’. This was done because existing pipelines are not likely to increase the cost of subsequent pipeline construction, as they already have established a right of way. A schematic of the Clip Features model is shown in Figure 10.

**Convert Features to Rasters – Level 1**

Figure 11 illustrates the model for converting the features into rasters. The features in the feature dataset created in the ‘Top Level’ model are iterated through. Each one has a new field added to the attribute. This field is used to assign the value 1 to each feature in each feature class. The feature classes are then converted to rasters using this the value field for the conversion. The rasters created have a value of one where a feature existed and zero everywhere else. These rasters have the same extent and cell size as specified by the user. The rasters that are no longer necessary for the analysis are then deleted.

**Generate Cost Surface – Level 1**

The cost surface is generated by using ‘Calculate Field’. There are four steps which sequentially see if the raster has a higher cost factor defined by Table 3 for every given cell. The steps start out by setting the value of a cell to 3 if it contains a road or railroad. This value is then set to 10 if the reclassed land cover raster is 10 (signifying a waterbody) or it contains a flowline or waterbody from the NHDPlus data. If there is a state park or a reclassed land cover value of 15 (developed land or wetland) the cell takes a value of 15. Finally, if it is coincident with a National Park the value is set to 30. These output from the model is a raster ‘Land Weight’, Figure 12.

**Excel to Point Sources – Level 1**

The source comma separated variable file is converted to an XY Event layer. As this is only temporary, the ‘Feature Class to Feature Class’ tool is implemented to make the points permanent, Figure 13.
Least Cost Paths – Level 1

The model, Figure 14, takes the land weights raster and adds 1 to every cell to ensure it has a base cost factor of 1. It also adds the cost factor for the slope to this, resulting in the ‘Total Weight’ raster, which defines the cost surface for the cost distance analysis. A raster catalog is created, which along with the cost surface raster and point source feature class, created in the ‘Excel to Point Sources’ model, is input to the ‘Iterate Sinks’ model.

Iterate Sinks – Level 2

This model iterates through the point feature class containing all of the sinks and calculates cost distance and cost backlink rasters for each one, using the previously calculated cost surface. These rasters are provided to the ‘Cost Path’ tool with the point source feature class. The least cost path is then determined between each source and each sink, considering the cost surface that was previously defined. Figure 5 shows that the cost distance raster is unique for each origin point.

The output cost path rasters are then stored in the raster catalog. This allows multiple raster datasets to be viewed as a single layer and also helpfully has the effect, in this case, of redefining the nodes across the multiple rasters so that there aren’t separate nodes assigned the same numeric value.

Each arc in the least cost path has a different value, with the value of 1 reserved for the starting cell and 2 reserved for arcs that are used by multiple routes. The costs given in the attribute table are for the total length from source to sink, and not for individual arcs. Hence, any arc with a value of 2 has a cost of 0 because it does not terminate at a defined destination. For a network optimization, the individual, not cumulative, costs of all the arcs are necessary.

The schematic in Figure 15 summarizes the model.

Node Network – Level 1
The model, Figure 16, converts the raster catalog back to a raster dataset. This is then converted to a polyline, with dangling lines of less than two times the cell size removed. These small, extraneous lines occur at the end of the network near the sources and sinks. The integrate tool smooths the path slightly, but is primarily used to remove small loops and extraneous polylines within the middle of the network. This tool integrates over two times the cell size also. The resulting polyline feature class has the field ‘Cost’ added to the attribute table.

**Node Costs Top – Level 1**

The purpose of this model, and its sub-model, is to determine the value of the cost distance raster at each node of the network, so that the cost of each arc can be determined. Figure 17 illustrates the iteration through the cost distance rasters generated in the model ‘Iterate Sinks’. At each iteration, a field is added to the network polyline feature class. The field header in each case is the same as the raster name of the current iteration. These data are then passed to the sub-model ‘Node Cost Features’.

**Node Costs Features – Level 2**

For each cost distance raster, the model calculates the cost of each arc in the output network, Figure 18. The model does this by iterating through the individual features in the output network and turning each one into its own feature class. These individual features consist of an arc and two end nodes. These end nodes, or vertices, are converted into points. The cost distance raster is sampled at these points and the range in cost values is then calculated. The attribute table field created in the parent model, ‘Node Network’, is then set equal to the range.

Unless, the arc is a common trunk-line resulting from the least cost path generation for multiple sinks, there will be a difference in the range in cost values for a given set of nodes for each different cost distance raster. This is due to the cost value for a particular cell in the cost distance raster only
being applicable for getting there, via the least cost path, from the specified input source location. Counter-intuitively, the correct value for the arc cost is the largest of all of the calculated ranges.

The output of the model is the node network with an attribute table containing unique node identifier, arc identifier, arc lengths and arc costs, *Figure 6*.

**Further Work - Model**

Once the implementation of an optimization code has been completed, the model should be revisited to add additional functionality and analysis:

- Inclusion of ports and docks as ship transport of CO₂ provides a low risk option for early stage, offshore CCS projects.
- Include existing pipelines in the cost factor determination. It is expected current rights of way will reduce the cost of building a new pipeline.
- Integrate the ArcMap model with the optimization software so that the network attributes will be directly exported into the optimization. Once the optimization has run, the optimized network will be displayed in ArcMap with arc line widths, and source and sink points scaled proportionally to the utilized capacity.

**Optimization**

A deterministic network model was created in IBM’s ILOG CPLEX Optimization Studio. The user defines the number of sources, sinks and nodes. The supply of CO₂ at each source node is specified, as well as the capacity of each sink and the total amount of CO₂ to be stored. This differs from a typical transportation problem in the fact that neither of the supply or demand has to be completely fulfilled.

The optimization uses a generic data file in which each arc is defined:

- Start node
- End node
The first three bullet points are determined directly from the ArcMap model. The optimization has been run on a simplified test case, but is yet to be trialed with the data from ArcMap.

**Conclusion**

ArcGIS and ModelBuilder provide an ideal workspace for creating a parametrized model to generate a node network with arc costs. This power of this lies in the ability to distribute the tools to other ArcGIS users, so that it can be used for analysis anywhere in the contiguous USA or even worldwide given suitable data availability. Furthermore, by simply changing the cost factor weightings in the ‘Generate Cost Surface’ model allows for other types of analysis. Such analyses could include natural gas transport to LNG terminals or Marcellus wastewater disposal from Pennsylvania to Ohio.
References

ArcGIS Elevation Server, NED 30m, retrieved from: http://landscape1.arcgis.com/arcgis


National Energy Technology Laboratory, 2015, NATCARB_Sources2013 (v1502). Retrieved from: https://edx.netl.doe.gov/dataset/natcarb-sources-v1502/resource_download/76df1e7d-ffc0-4ab6-8806-9003239f8be6


APPENDIX A
Figure 1, High Purity CO₂ Sources in Houston
Figure 2, High Purity CO₂ Sources in Port Arthur
Figure 3, User inputs into the ArcMap Model
Figure 4, Hierarchical Layout of ArcMap Model, Grey = Level 0, Blue = Level 1 and Green = Level 2
Figure 5, Showing the Cost Distance Raster from a) Sink 1 and b) Sink 2
Network Map of Potential Paths Between Sources and Sinks

Figure 6, Map of Node Network Paths and Costs
APPENDIX B – MODEL DESCRIPTIONS AND INPUTS
Figure 7, Top Level Model

Figure 8, Extract Rasters Model
Table 1, Reclassifying from Land Cover Values (United States Geological Survey, 2011) to Cost Factors (MIT, 2006)

<table>
<thead>
<tr>
<th>NLCD Value</th>
<th>Description</th>
<th>Cost Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Open Water</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>Perennial Ice/Snow</td>
<td>10</td>
</tr>
<tr>
<td>21</td>
<td>Developed Open Space</td>
<td>15</td>
</tr>
<tr>
<td>22</td>
<td>Developed Low Intensity</td>
<td>15</td>
</tr>
<tr>
<td>23</td>
<td>Developed Medium Intensity</td>
<td>15</td>
</tr>
<tr>
<td>24</td>
<td>Developed High Intensity</td>
<td>15</td>
</tr>
<tr>
<td>31</td>
<td>Barren Land (Rock/Clay/Sand)</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>Deciduous Forest</td>
<td>0</td>
</tr>
<tr>
<td>42</td>
<td>Evergreen Forest</td>
<td>0</td>
</tr>
<tr>
<td>43</td>
<td>Mixed Forest</td>
<td>0</td>
</tr>
<tr>
<td>51</td>
<td>Dwarf Scrub</td>
<td>0</td>
</tr>
<tr>
<td>52</td>
<td>Shrub/Scrub</td>
<td>0</td>
</tr>
<tr>
<td>71</td>
<td>Grassland/Herbaceous</td>
<td>0</td>
</tr>
<tr>
<td>72</td>
<td>Sedge/Herbaceous</td>
<td>0</td>
</tr>
<tr>
<td>73</td>
<td>Lichens</td>
<td>0</td>
</tr>
<tr>
<td>74</td>
<td>Moss</td>
<td>0</td>
</tr>
<tr>
<td>81</td>
<td>Pasture/Hay</td>
<td>0</td>
</tr>
<tr>
<td>82</td>
<td>Cultivated Crops</td>
<td>0</td>
</tr>
<tr>
<td>90</td>
<td>Woody Wetlands</td>
<td>15</td>
</tr>
<tr>
<td>95</td>
<td>Emergent Herbaceous Wetlands</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 2, Reclassifying from slope to Cost Factors (MIT, 2006)

<table>
<thead>
<tr>
<th>Slope</th>
<th>Cost Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-20%</td>
<td>0.1</td>
</tr>
<tr>
<td>20-30%</td>
<td>0.4</td>
</tr>
<tr>
<td>&gt;30%</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Table 3, Carbon Dioxide Transportation Cost Factors (MIT, 2006)

<table>
<thead>
<tr>
<th>Description</th>
<th>Cost Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossing</td>
<td></td>
</tr>
<tr>
<td>Road</td>
<td>3</td>
</tr>
<tr>
<td>Railroad</td>
<td>3</td>
</tr>
<tr>
<td>Waterway</td>
<td>10</td>
</tr>
<tr>
<td>Protected Area</td>
<td></td>
</tr>
<tr>
<td>Populated Area</td>
<td>15</td>
</tr>
<tr>
<td>Wetland</td>
<td>15</td>
</tr>
<tr>
<td>National Park</td>
<td>30</td>
</tr>
<tr>
<td>State Park</td>
<td>15</td>
</tr>
</tbody>
</table>
Figure 9, Resample Model

Figure 10, Clip Features
Figure 11, Convert Features to Rasters

Figure 12, Cost Surface Calculation Model
Figure 13, Excel to Point Source Model

Figure 14, Least Cost Paths Model
Figure 15, Iterate Sinks Model

Figure 16, Node Network Model
Figure 17, Node Costs Top Model
Figure 18, Node Cost Features Model
Peter Tutton pmt553

```plaintext
int Nnodes = ...; // Number of nodes
range nodes = 1..Nnodes; // Range of nodes
int Nsources = ...; // Number of sources
range sources = 1..Nsources; // Range of sources
int Nsinks = ...; // Number of sinks
range sinks = 1..Nsinks; // Range of sinks

int supply[sources] = ...;
int demand = 5;

// Create a record to hold information about each arc
tuple sourcearc {
    key int fromS;
    key int toN;
    float cost;
    float bound;
}
tuple nodearc {
    key int fromN;
    key int toSink;
    float cost;
    float bound;
}
{sourcearc} Sarcs = ...;
{nodearc} Narcs = ...;

dvar float+ SourceO[s in Sarcs] in 0 .. s.bound;
dvar float+ SinkI[n in Narcs] in 0 .. n.bound;
dexpr float TotalFlow = sum (s in Sarcs) s.cost * SourceO[s] + sum (n in Narcs) n.cost * SinkI[n];
minimize TotalFlow;
subject to {
    // Preserve flows at each node. Note the use of slicing
    forall (x in sources)
        ctSourceFlow:
            sum (<x,t,c,ub> in Sarcs) SourceO[<x,t,c,ub>] <= supply[x];
    forall (n in nodes)
        ctNodeFlow:
            sum (<f,n,c,ub> in Sarcs) SourceO[<f,n,c,ub>] - sum (<n,g,c,ub> in Narcs) SinkI[<n,g,c,ub>] == 0;
        ctSinkFlow:
            sum(<f,t, c, ub> in Narcs) SinkI[<f,t,c,ub>] == demand;
}

execute  {
}
```

Figure 19, Mod file for executing simple network optimization with a total demand of 5
Nnodes = 2;
Nsources = 2;
Nsinks = 2;

supply = [2 5];

Sarcs = {
    <1, 1, 13, 2>
    <1, 2, 5, 1>
    <2, 1, 15, 4>
    <2, 2, 8, 4>};

Narcs = {
    <1, 1, 1, 10>
    <1, 2, 2, 1>
    <2, 1, 3, 10>
    <2, 2, 8, 10>};

Figure 20, Data file for optimization specifying 2 sinks, 2 nodes and 2 source with a supply of 2 and 5