International Coal Market Module

Component Design Report

OnLocation, Inc.

4/14/2017

Prepared For

U.S. Energy Information Administration

Page left intentionally blank

Contents

List of Figures

Executive Summary

The U.S. Energy Information Administration (EIA) is currently seeking to develop a new International Coal Market Module (ICMM) as part of the World Energy Projection System Plus (WEPS+). Towards this end, EIA has commissioned this Component Design Report (CDR) from OnLocation to give recommendations for the design of the new coal model. The main purpose of the ICMM will be to project regional coal prices and coal supply, as well as global coal trade flows. There are several key challenges in modeling the international coal market:

- Modeling realistic changes in regional coal production over time.
- Representing evolving patterns of interregional coal trade in response to changing patterns of consumption and production.
- Projecting regional delivered prices to each consuming sector over time for various coal types.

This CDR describes in detail an approach to the ICMM that will allow EIA to meet these challenges and enable the model to be readily adapted based on new information. The approach described in this report is designed to be data-driven with respect to model regionality and time horizon. This CDR addresses the major areas of interest stated by EIA, including:

- Major drivers of the international coal market
- Seaborne and inland transport of coal
- Regionality
- Representation of coal production
- Coal quality
- Data sources
- Software framework

In summary, the proposed modeling approach is a network optimization linking supply and demand nodes where coal production opportunities are expressed as supply steps that are econometrically derived. Coal reserves and production are represented by coal type at the national or sub-national level as appropriate. Inland and international transportation costs are explicitly represented. Expansion of production and transportation capacities are endogenously projected based on economic conditions. Delivered coal prices are based on marginal costs of supply (production and transport) plus relevant taxes and fees.

The data will be stored at the most elemental level available and then aggregated into regions for modeling purposes. A geographic information system approach to managing the spatial data in the ICMM is recommended (but not required.)

1 Overview

The International Coal Market Module (ICMM) is a key module of the World Energy Projection System Plus (WEPS+) due to the evolving international approach to climate change that is impacting coal use and transport. The simple current approach needs a major revision to reflect the increasing importance of coal in forecasts of energy flows globally. The U.S. Energy Information Administration (EIA) is also reexamining other models to make them more responsive to technology and policy changes that will influence global energy production and consumption. The new international coal model will need to provide key inputs into these models and be responsive to changing environmental and economic policies. EIA has identified a list of requirements that comprise the scope and overall considerations that must be addressed by the new model. This document outlines the approach OnLocation recommends as the design for a new international coal model.

EIA faces a number of challenges in developing the new ICMM. Among the more critical are:

- Modeling realistic changes in regional coal production over time.
- Representing evolving patterns of interregional coal trade in response to changing patterns of consumption and production.
- Projecting regional delivered prices to each consuming sector over time for various coal types.

Any new model that EIA develops should be able to address each of these in a way that is flexible in the regional representation and robust across a wide range of analysis scenarios. This Component Design Report (CDR) describes in detail an approach to the ICMM that will allow EIA to meet these challenges and enable the model to be readily adapted based on new information as it becomes available (production patterns, consumption patterns, and policies).

Figure 1 provides an overview of the components of the design that will be described in detail in the CDR. The model design includes a front-end preprocessor that manages the data and is used in the regional definition formulation. The core of the model includes a linear programming (LP) optimization approach. The back-end processes data for integration with the rest of WEPS+.

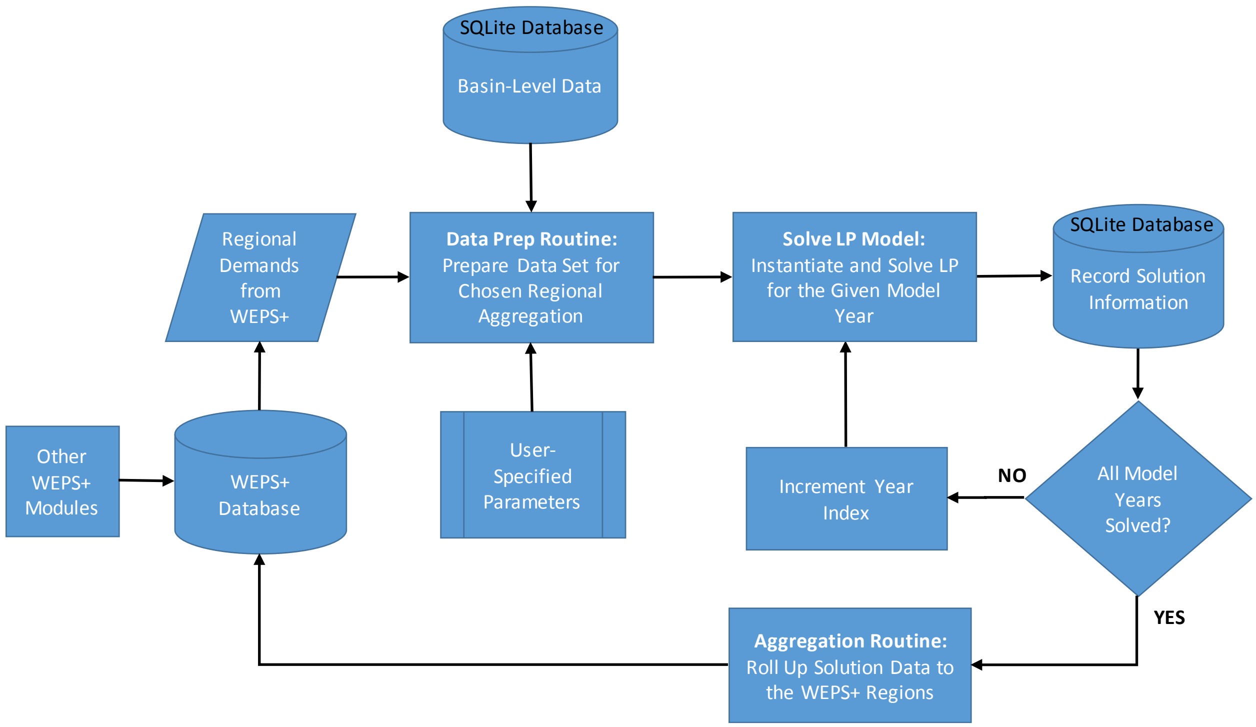

Figure 1: Block Diagram of ICMM Components

The design of the modeling approach is flexible and responsive to change. As new and more information becomes available in various regions of the world, the design allows for this information to be readily incorporated into the definition of the model objects. By combining a knowledge-based process with a flexible solution choice, the design inherently allows the variety of regional circumstances to be captured in the coal market forecast.

The design of the modeling approach will allow the resultant model and database framework to be maintained, modified, and extended by EIA, using software currently in use at EIA and/or publicly available. The design uses SQLite (which is currently used by the WEPS+ team) to contain all of the data (raw data can still be stored in Excel files or in comma-separated values, CSV, files). The main LP model will be written and solved in the AIMMS mathematical modeling language (Advanced Interactive Multidimensional Modeling System), and Python will be used for data manipulation and interfacing with WEPS+ or other modeling frameworks.

This design addresses all of EIA's needs, it is flexible and responsive to change, and it can be maintained by EIA.

1.1 Background and Key Drivers

Despite the widespread perception that the coal industry is in decline, the reality is that coal will likely be the dominant fuel for both electricity production and steel production for many years to come. Coal is abundant and cheap, easy to transport, and does not have the geopolitical concerns found in the oil and gas industries. Because of this, coal will be slow to displace in the near future as an energy source despite having several well-known environmental downsides.

There are several key drivers in the global coal market which impact the forecasts for coal demand, supply, and pricing which we will outline here.

1.1.1 China

The main story in the global coal market will continue to be China for many years to come. China accounted for approximately 46% of global coal production and 50% of global demand in 2015 according to the International Energy Agency (IEA) *2016 Medium-Term Coal Market Report*. While the demand for coal in China is expected to level off in the near future, forecasts of future coal markets will be highly dependent on assumptions regarding the macroeconomic conditions and policy regime in China because of its dominant production and consumption shares.

Figure 2 depicts major importers, exporters, and trade flows of steam coal in 2004 and 2014.[1] This figure clearly shows the emergence of China and India as dominant players in the global import market. China moved from being a net exporter of coal in 2004 to being the largest importer in 2014. In the prior two decades before 2010, Japan and South Korea were the largest importers of steam coal. In the time since, China and India have surpassed Japan as the largest importers. In general, the global coal trade shift towards Asia is expected to continue.[2]

---

1 UN COMTRADE Database (https://comtrade.un.org/data/).

2 From the EIA *International Energy Outlook 2016* (U.S. Energy Information Administration, 2016).

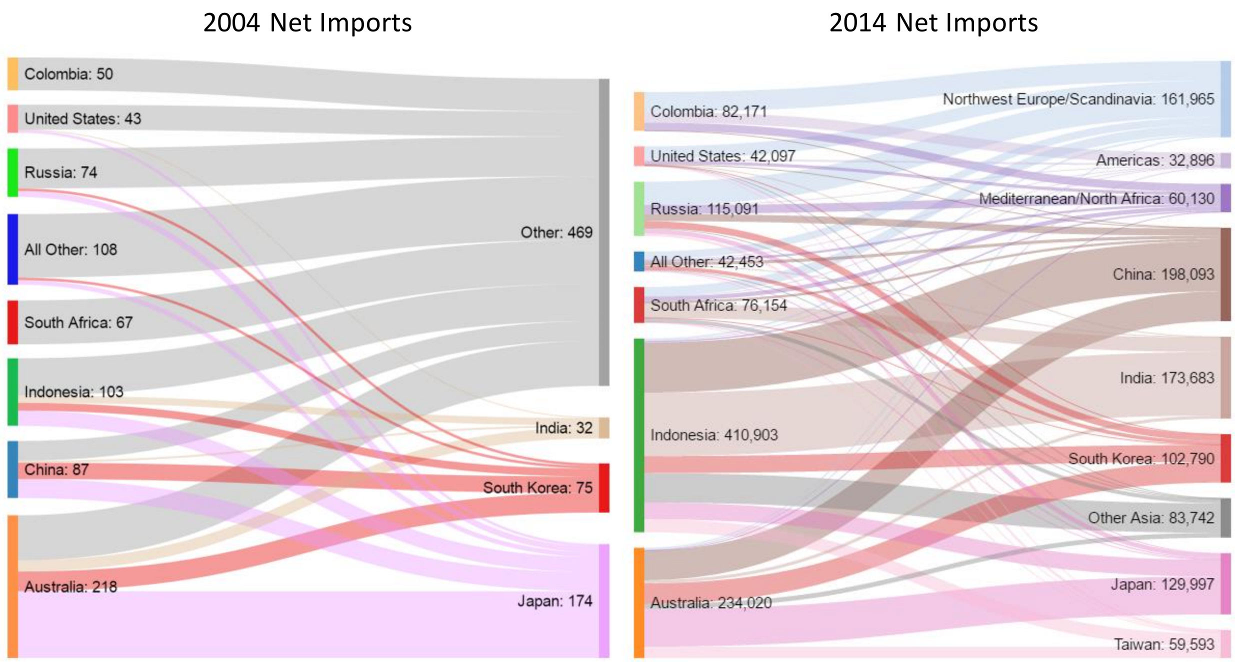

Figure 2: Net Coal Trade Flows for Years 2004 and 2014

1.1.2 ASEAN Countries

Coal demand in India and the Association of Southeast Asian Nations (ASEAN) member states is expected to continue experiencing strong growth. While the scale is not expected to be comparable to the recent increases in China, which has been adding approximately 200 MW of coal generating capacity *daily* over the past decade, demand growth in these counties will still be a major driver of global coal markets in the next several decades.

While India has significant coal reserves, India has become a coal importer as demand growth has exceeded supply expansion. The government has set ambitious targets for increasing production, but it is not clear that production and rail delivery investments will be sufficient to meet the goals.[3]

1.1.3 Renewable Energy and Climate Policy

Demand for coal has been decreasing in North America and Europe due to climate policies and technology developments in renewable energy. According to EIA, coal consumption for electricity generation in the United States has decreased from 1041 million short tons in 2007 to 738 million short tons in 2015, a 29% reduction. This is partly due to cheap natural gas

---

[3] From the EIA International Energy Outlook 2016 (U.S. Energy Information Administration, 2016)

supplies stemming from the U.S. fracking revolution, but is also due to the investment in natural gas generation being seen as less risky because of its smaller carbon footprint. The assumptions regarding the pace of technological advancement in renewable generation technologies as well as future climate change mitigation policies for each country will have a large impact on the future global coal market.

One example of a global climate policy assumption that would have a huge impact on the global coal market is the recent Paris Agreement on climate change. The extent to which signatory governments enforce the CO2 emission limits outlined in the agreement will have a big impact on the global coal market, as will development of enabling technologies such as carbon capture and storage (CCS). Meeting the Paris targets without an ability to capture CO2 emissions would require an almost complete elimination of coal as an energy resource.[4]

1.2 Statement of Model Purpose

The purpose of the ICMM is to project regional wholesale and retail coal prices, regional coal production volumes, and the trade flows of multiple coal types over the forecasting horizon subject to a set of regional demands for coal provided by other WEPS+ modules. It is currently envisioned that the ICMM will be integrated into the WEPS+ framework as a replacement for the current coal module.

The ICMM must be responsive to a wide range of scenarios that alter the coal production volumes, global coal trade, and coal production costs based on varying assumptions. Some of these assumption could originate in other WEPS+ modules. For example, changes in regional coal demands will be driven predominantly by assumptions made in the WEPS+ International Electricity Market Module (IEMM). Other assumptions could be directly related to coal production and transportation costs that would originate in the ICMM input data.

1.3 Flexible Design

The modeling approach to the ICMM we are proposing is built around flexibility, transparency, and a responsive software development process. While there is a myriad of approaches to

---

[4] The carbon budget available for limiting temperature increases to less than 1.5°C with a likelihood of 66% is estimated by the Intergovernmental Panel on Climate Change (IPCC) *Climate Change 2014 Synthesis Report* (IPCC, 2014) to be 400 gigatons of carbon dioxide (GtCO2) starting in 2011. Total global CO2 emissions were 175 GtCO2 from 2011 to 2015, which leaves a remaining carbon budget of 225 GtCO2 *from 2015 on* if global temperature increases are to be kept to under 1.5°C. According to the EIA *International Energy Outlook 2016*, the annual global CO2 from burning coal in 2012 was approximately 15 GtCO2, so the remaining budget of 225 GtCO2 equates to approximately 15 years of burning coal at our current rates. And this does not even take into account the CO2 emissions from other sources.

addressing this modeling problem, we believe the combination of the auto-generation of input data sets from raw data and the ability to easily change model regionality will provide the benefits outlined in the following sections. Furthermore, the optional addition of a geographic information system (GIS) for the visualization and maintenance of ICMM spatial input data could provide an added level of efficiency.

1.3.1 Ease of Use and Maintenance

Our recommended approach to data management in the ICMM is to store data that has a spatial component in a GIS database. A GIS would enable the data sets to be analyzed by modern geographic information systems that allow the user to inspect, verify, and validate the data sets. Updating the inputs to the model is then an exercise of updating and maintaining the GIS data, which may be done graphically. Furthermore, this approach lends itself to distributed upkeep by allowing regional/country experts to be responsible for the content of each data set, while modeling experts manage their input and use within the ICMM. Finally, to the extent applicable, Application Program Interfaces (APIs) can be employed to periodically refresh the data sets automatically from online sources.[5] The use of a GIS for the ICMM, as well as alternatives, are discussed in Section 3.5.

1.3.2 Flexible LP Modeling Framework

The ICMM LP will be written in a data-driven framework so that changes to key problem dimensions such as regionality and time horizon require no changes to the code describing the LP variables, constraints, and objective function. This will allow the users to focus on the preparation of alternate data sets and analysis of model results instead of spending time making technical changes to the core LP model.

Taking this approach, the model will support evolutionary development, and continuous improvement, and it will encourage rapid and flexible response to change.

1.3.3 Regional Disaggregation

As described more fully below, our approach is centered on the flexible definition of both supply and demand sub-regions that decompose the WEPS+ aggregate regions into logical abstract objects representing shares of WEPS+ regional coal demands to be passed to the LP solver along with regional supply regions and the potential links between them. In its most simple form, the supply and demand regions can be left aggregated at the WEPS+ regional level

---

[5] The EIA Open Data API found at https://www.eia.gov/opendata/ is one example of an online data source around which automated data updates could be created.

and sent to the solver directly. Note that this formulation works no matter what level of data detail is defined as the chosen aggregation level for a particular region can reflect the level of data detail that currently exists. The aggregation can be made to be finer as more detailed information becomes available.

1.3.4 Responsive Software Design

Our approach is to use off-the-shelf core software (e.g., Python and AIMMS), in conjunction with GIS software and SQLite database software to build out the ICMM. We will separate the raw data from the model inputs so that alternative sets of model inputs may be created easily. The solution approach will stand by itself and be testable outside the full implementation. The ICMM specific report writing and output databases will be separated into their own module. By taking this approach, EIA will be able to adapt and evolve the design as new data and concepts for solutions become available. Further, due to the flexible framework, EIA will be able to test alternative levels of aggregation. This approach will allow development and early delivery of core pieces for testing and evaluation and for continuous improvement; it also encourages rapid and flexible response to change both during the development cycle and long term.

1.4 Comparisons to Other Modeling Approaches

A brief review of other approaches provides useful context before the proposed new design for the ICMM is described. In particular, the Coal Market Module (CMM) of the National Energy Modeling System (NEMS), with its domestic and international component models, is well known to EIA and provides a good reference point. The recent work by Hellerworx builds on that methodology. The COALMOD-World model uses a different framework but shares some common features with the CMM.

1.4.1 Coal Market Module of NEMS[6]

The NEMS coal model includes domestic production, domestic transportation, and an international submodule that uses global supply and demand to determine the price and quantity of U.S. coal imports and exports. The production model uses a regression approach to estimate supply curves relating price to quantity produced by region, mine type, and coal type. Coal types include four thermal types (bituminous, subbituminous, lignite, and metallurgical) and three sulfur grades of coal for surface and underground mining for a total of 12 types (not all combinations exist). Inputs to the model include user-specified inputs for the base year that include capacity utilization at mines, productive capacity, mine mouth coal prices, miner wages, labor productivity, cost of mining equipment, and the price of electricity. User-specified inputs

---

[6] See (US Energy Information Administration, 2014).

for forecast years include annual growth rates for labor productivity and wages, and annual producer price indices for the other non-fuel production costs, such as the cost of mining machinery and equipment, iron and steel, and explosives. Fuel prices and the real interest rate are provided by other modules of NEMS, as are prior year values for coal production and prices.

Domestic distribution is determined through LP algorithm that considers transportation costs along with supply costs in minimizing the cost of delivered coal to each region. These transport costs vary over time based on a time trend and changes in labor and fuel costs, as well as the cost of capital. The model also takes into account existing contracts. The LP includes environmental constraints that limit various criteria pollutant emissions. The model provides coal prices, sulfur content, mercury content, and sulfur dioxide (SO_2) and mercury allowance prices (if applicable) to the EMM. The demands from the EMM are for British thermal units (Btu) by generator type. Generator type is focused on the emissions based on the configuration of pollution and toxin controls. Using this information, the CMM determines the fuel switching among coal types that is necessary for environmental compliance.

Transportation costs between supply and demand nodes by sector are inferred from historical data of delivered and mine mouth prices. This technique is used because coal rail transportation rates in the United States are viewed as not being fully competitive and hence are not based exclusively on cost or distance. A second set of higher rates is used in some instances for prices to the electricity sector for expanding transportation volumes. Both sets of transportation rates are modified over time using regression equations, with separate equations for coal originating in the eastern regions versus western regions. The indexes depend on a variety of factors including railroad productivity, the user cost of capital for railroad equipment, the national average diesel fuel price, gross capital expenditures for Class I railroads, and western share of national coal consumption.

The international trade module of the CMM projects U.S. coal exports and imports in the context of world coal trade by using an LP approach to determine the flows that minimize coal delivery cost for steam and metallurgical coals among a set of 17 coal export and 20 coal import regions. The key user-specified inputs for the non-U.S. regions include coal import demands, coal export curves, transportation costs, and various constraints. Several constraints are imposed including maximum export capacity, maximum shares of import and export regions from other regions to represent diversity of supplies so that one region does not rely on only one trade partner, sulfur and mercury limits if active in the United States, and existing contract quantities.

The step-function coal export curves contain information about the quality of coal (sulfur, mercury, CO_2, and heat content), the base year export capacity, the base year export free-on-board (FOB) price, and then price/capacity pairs for future years in 5 year steps. There is also a scalar that allows the prices to be adjusted over time relative to the U.S. coal export price based on expected differences in productivity. A supply curve may have multiple price tranches, but they must all have the same coal quality. The cost associated with each quantity of coal available for export includes (1) mining costs; (2) representative coal preparation costs, which may vary according to export region, coal type, and end-use market; and (3) inland transportation costs (prior to export). Because the underlying data are sparse, significant analyst judgment may be necessary in the development of these curves. Analysts may be required to fill in missing data based on educated guesswork based on literature research and inference from known data.

International coal transportation rates are developed for all feasible combinations of import and export regions for thermal and coking coal. These rates are developed as a function of route distance and the applicable maximum vessel size.

1.4.2 Hellerworx, Inc., Australian Pilot Study[7]

Hellerworx was contracted by EIA to develop a methodology for creating the international export curves used in the CMM for projecting U.S. coal imports and exports and to review and improve the econometric-based method used to project international freight costs. A coal supply curve was estimated for Australia as a pilot study. As part of the study, Hellerworx also identified potential data sources for developing supply curves for other countries.

A regression model was estimated that relates coal export volumes by type (thermal and coking) to export prices. The most statistically relevant independent variable was determined to be labor productivity. Other variables were tested, including diesel prices and wages, but were found not to statistically contribute in the case of Australia. Development of equations using a disaggregation of sub-regions within New South Wales and mine type (underground and surface) did not produce better statistical results than the aggregate fits for New South Wales and Queensland. While inland transportation costs were estimated, it is not clear how they were used in the supply curve estimation process.

An engineering cost model was developed to estimate international freight costs for two vessel types as a replacement to the relatively simple equations currently used in the CMM. A fixed portion in US dollars per metric ton ($/MT) of the cost is composed of port costs, daily hire rates, number of port days, cost of diesel use in port, and cargo size. The variable portion in US dollars per metric ton-mile ($/MT/mile) is a function of the price of bunker fuel, fuel use per

---

[7] See (Hellerworx, Inc., 2016)

day, daily hire rates, miles traveled per day, and cargo size. The model was then calibrated to match average rates, as the model tended to underestimate costs for short-distance trips and overstate for long trips. New build costs for each vessel type were also used to establish long-term equilibrium rates.

1.4.3 COALMOD-World[8]

COALMOD-World is a multi-period model of coal supply, demand, and international trade that includes investments in coal production and transportation capacities. The model assumes profit-maximizing players who optimize their expected and discounted profit over the total model horizon. The model represents asymmetric consumption and production nodes with more than one consumption and/or production node per country for large countries that have high demand or production and geographically separated production areas. In addition, there are export nodes that are linked to producer nodes. Producers are assumed to maximize profit given production capacity constraints and production and the inland transport costs. They can sell coal either to local demand nodes or to linked exporters. In addition, producers can invest in expanding production capacities and in transport capacities to local demand or to the exporter. Exporters also maximize profits as defined by “the revenue from sales net of the costs of purchasing the coal at the FOB price from the producer, the costs of operating the export terminal, the costs of transport (shipping) to the final market and finally the potential costs of investing in additional export capacity.”[9]

Linear production cost curves are estimated using estimates for low and high average costs with the low cost representing the y-intercept at zero production. The high cost is equated to the maximum annual production capacity. Assumptions are made about the mine depletion rates, called “mortality” rates, that shift the curves upward over time. Investment opportunities are characterized by an investment cost in dollars per ton of annual capacity addition and are assumed to be restricted per period. Total reserves are static. The coal of each production node is assigned an average heat content.

Land transportation costs from production to consumer and export nodes are assumed to be constant over time, but the capacities can be expanded by investments. International freight costs are based on a linear regression equation as a function of travel distance, and these costs are constant over time. Port costs and capacities are assessed for each export node, along with investment costs for expansion. As with production expansion, maximum expansion rates are set for port capacities.

---

8 (Holz, Haftendorn, Mendelevitch, & von Hirschhausen, 2016)
9 (Holz, Haftendorn, Mendelevitch, & von Hirschhausen, 2016), page 10.

2 Input and Output Requirements

2.1 Input Requirements

The new International Coal Market Module (ICMM) will require information from other models within the World Energy Projection System Plus (WEPS+) as well as external data. Section 3 provides a more detailed assessment of the data needs and their potential sources.

2.1.1 Passed from Other Modules

The primary data required from the other WEPS+ models (passed via a binary restart file, as in the National Energy Modeling System, NEMS) are sectoral coal demands. At present these demands are only distiguished by WEPS+ region, but it is anticipated that further detail by coal type, coal quality, and additional regionality will be available in the future. The ICMM needs to be adaptable to accept these demands at whatever level of granularity is defined in the other models as they continue to evolve. The largest coal demand source by far is the electricity sector, so communication with the International Energy Market Module (IEMM) will be especially important.

The macroeconomic model will provide financial and economic inputs such as interest rates, general wage rates, and potentially exchange rates.

Delivered fuel prices by sector and each of the 16 WEPS+ regions are available from the other supply models. The ICMM will use delivered prices to the transportation sector for distillate and residual oil in the computation of coal freight costs.

2.1.2 Data Coming from Exogenous Input Files

The coal model will rely on considerable additional exogenous inputs related to coal production, inland transportation, and international transportation. Historical time series may be stored for convenience in addition to model parameters necessary for the projections. These will be stored in a flexible database as described in Section 5.3.

2.2 Data Assessment and Sources

In this section we step through the broad categories of information that are necessary to forecast global coal production, trade flows, and prices. The data should be gathered at the lowest level of geographic aggregation as possible and can then be processed within the model to the appropriate region definitions in use.

2.2.1 Coal Production, Capacity, and Production Costs

Historical data on coal production, coal mining employment, production costs, mining capacity, and capacity expansion costs are needed at the finest geographic and coal quality and mine type detail available. Because global coal production is very concentrated in just a few countries, this is not for the most part as daunting a prospect as it seems, although data on coal quality is likely to be difficult to find. Where data are missing, data will need to be imputed, in many cases by using average values available across a more aggregate data set.

Potential data sources include:

International Energy Agency (IEA) country level production data contain coal production by coal type and country.

IEA *Medium-Term Coal Market Reports* include multiple charts that display coal supply costs and prices that source the Wood Mackenzie private coal database and *McCloskey Coal Report*s.

IHS *China Coal Production and Cost Outlook* provides raw and thermal coal production by production unit within China for a few historical years and then projections. Each production unit is identified by province and coal type (steam or metallurgical) and by mining method (surface or underground). China's mine-gate cost of raw thermal coal (value-added tax included) is also provided from 2011 to 2040.

IHS single country data sets (Indonesia example) contain a couple of years of historical data on production, exports, and domestic production by mine/company.

World Energy Council, *World Energy Resources: 2016* is a publicly available document with information on coal resources, reserves, and production that is compiled from multiple data sources, including *BP Statistical Review of World Energy 2016*, BGR Energy, WEC Energy, and IEA, and from country-specific information pulled from a variety of sources.

Single-country reports are available from various government and commercial sources. A sample of these is from the Hellerworx report:

- <u>Australia:</u> New South Wales data compiled by Coal Services Pty Limited and available on a subscription basis; Queensland government data website
- <u>Indonesia:</u> In addition to IHS, local data provider CoalInfo may have data
- <u>Poland:</u> Government statistical site with data on production, exports, and export prices
- <u>Russia:</u> *Russian Coal Journal* has some data on production, productivity, and export prices
- <u>South Africa:</u> Mostly IHS, but also a few other sources

- Canada: *Statistics Canada Coal Monthly* has monthly production data
- Colombia: Productivity data could be developed using data from Cerrejon, which is the largest coal producer

The **COALMOD-World** documentation provides an extensive bibliography with data sources for:

- Coal production costs
- Information on coal resources, reserves, and production for hard coal and lignite for top 20 countries and aggregate regions
- Average heat content of coal by area
- Coal production and export terminal capacity expansion costs

2.2.2 Inland Transportation Costs and Capacities

Costs to move coal from mine to either domestic consumers or ports for exports can be derived either from pricing information or built up from rail rates. For example, the Coal Market Module (CMM) uses U.S. transport costs derived from the historical difference between delivered and mine mouth prices. For Australia, Hellerworx estimated transportation costs using a combination of proprietary and public data, but indicated that the costs could be updated using public information of rail rates. Transport costs in COALMOD were based on a 2007 data source.[10] If older data are used for completeness, that data should be updated using rail rates or other information from more current sources.

2.2.3 International Transportation Costs

The components for international shipping costs include port costs, days in port, ship and labor rates (could be as a daily hire rate), fuel consumption rates, and fuel prices. The Hellerworx report lays out costs for these components. As discussed earlier, fuel prices would be used from the supply side of WEPS+. Existing port capacities and expansion costs in COALMOD were estimated using data from several sources, including IEA, the Organization for Economic Cooperation and Development (OECD), and several research papers. EIA has access to estimates of throughput by port (22 ports in 8 countries) from IHS. Historical freight rates between ports since 2001 by ship size are also available from IHS.

[10] (Baruya, 2007)

2.2.4 Pricing Information

Retail coal prices comprise delivered prices marked up by excise and value-added taxes that may vary by coal type and sector.

Taxes on goods and services (or consumption taxes — value-added tax, "VAT", sales taxes, and other, similar taxes) ultimately impact the price of coal paid by the end user. Depending upon the type of tax, consumption taxes can also impact the effective price of imported and, in some cases, exported goods (i.e., energy imports and exports). Data on consumption taxes are publicly available and are tracked and published by a number of organizations including:

World Trade Organization, World Tariff Profiles, is a publicly available data set of duties applied to goods in services by country and tariff type.
https://www.wto.org/english/tratop_e/tariffs_e/tariff_data_e.htm

The World Bank, Taxes on Goods and Services, is a publicly available data set that tracks current and historical tariff data. http://data.worldbank.org/indicator/GC.TAX.GSRV.VA.ZS

2.2.5 Discount Rates and Risk Premiums

Discount rates and risk premiums dictate the cost of investment. The discount rates are typically set by the central bank of a country and represent the cost of lending between banks. Risk premiums represent the relative likelihood of an investor receiving a certain return on investment and are applied to discount rates to determine the cost of money borrowed. For the ICMM, discount rates and risk premiums for a region or country will be largely influenced by the sovereign credit rating of the country or member countries. Sovereign credit ratings are indicators of the relative economic and political risk of a country. Countries with good credit ratings (investment grade) have access to more funding sources at lower costs. For developing nations, a stronger credit rating also increases the attractiveness of foreign direct investment. Conversely, non-investment grade countries (i.e., those with poor credit ratings) are limited in their access to funding sources and incur greater costs on debt. Three primary credit rating agencies (Standard and Poor's, Moody's, and Fitch) evaluate the credit worthiness of countries and periodically publish their findings.

Central bank discount rates are published by a number of sources including:

Trading Economics, *Country Interest Rates*, which provides publicly available data on several economic indexes, including country (central bank) interest rates as well as sovereign credit ratings from the three major credit agencies.
http://www.tradingeconomics.com/forecast/interest-rate

Central Bank Rates, *Worldwide Central Bank Rates,* publicly available listing of central bank rates. http://www.cbrates.com/

U.S. Central Intelligence Agency, *World Fact Book: Central Bank Discount Rate,* publicly available publication of the annualized rate charged by a country's central bank to commercial, depository banks for loans to meet temporary shortages of funds. https://www.cia.gov/library/publications/the-world-factbook/rankorder/2207rank.html

2.3 Outputs

The ICMM will produce the following outputs at a minimum for each model solution year:

- Coal production by coal type and region
- Coal prices by coal type, and region
    - Mine-mouth prices
    - Transportation cost component
    - Domestic consumption prices by sector
    - Export prices
    - Import prices
- Inter-regional and intra-regional coal flows by coal type and transportation mode
- Coal production capacity expansion by coal type and region and associated investment expenditures
- Transportation infrastructure capacity expansion by region and associated investment expenditures

Based on the overall WEPS+ model requirements, selected model outputs will then be communicated to the other models for use as inputs for other modules and reporting.

2.3.1 Passed Back to Other Modules

Information passed back to the other WEPS+ models via the restart file includes, for each model year:

- Coal prices by coal type, and region (on a US dollar per Btu basis)
    - Mine-mouth prices
    - Transportation cost component
    - Domestic consumption prices by sector
    - Export prices
    - Import prices
- Supply elasticities used by demand and transformation models to improve convergence as the model iterates between modules

- Inter-regional and intra-regional coal flows by transportation mode
- Coal production by coal type and region (coal quantities will be transferred to other WEPS+ models in heat content units at the required regional aggregation level)

These reporting items are not restricted to being at the 16 WEPS+ regional level; they could include more detailed aggregations which may be used by other modules such as the IEMM.

Items provided in the restart file are available to graphical display system, GrafNEM.

2.3.2 Additional Coal Market Reports

Additional detailed reports relating to coal production and transportation will also be produced:

- Coal production capacity expansion by coal type and region and associated investment expenditures
- Transportation infrastructure capacity expansion by region and associated investment expenditures

2.3.3 Aggregating to WEPS+ Regions

Once retail prices are established for the ICMM regions, average prices will be constructed for the WEPS+ 16 regions. These will be weighted averages using the same assumptions about demands in the ICMM regions relative to the WEPS+ regions. Coal production and transportation will be aggregated at the WEPS+ region level and elsewhere reporting at the sub-region level to allow verification and validation.

It is important to note here that the above aggregation is for WEPS+ as it currently stands today. The ICMM will likely be operating at a finer geographic level than the 16 WEPS+ international regions. To the extent that other models, such as the IEMM, are operating with greater regional resolution, more granular output from ICMM could be stored in the restart file for IEMM's use. Furthermore, while WEPS+ currently only considers a single conglomerate coal type, much finer detail about coal quality for various sources may be accounted for in the ICMM once other models are prepared to give coal demands broken out by coal quality. There is more discussion regarding the modeling of coal qualities from different sources in Sections 3.2 and 3.3.

3 Classification Plans

3.1 Regional Aggregation

Coal production is highly concentrated, with the top 15 countries in 2015 accounting for 96 percent of production according to International Energy Agency (IEA) statistics (see Figure 3).[11] China alone produced 46 percent of the world's coal, and the top 5 countries supplied almost 80 percent. Therefore, we anticipate that several production regions will be represented in these high producing countries while some of the smallest producing countries will be aggregated together (for example, all of Africa except South Africa could be a single region). As described earlier, the input database will contain data at the lowest level of geographical aggregation possible that can then be combined into the regions selected by the user.

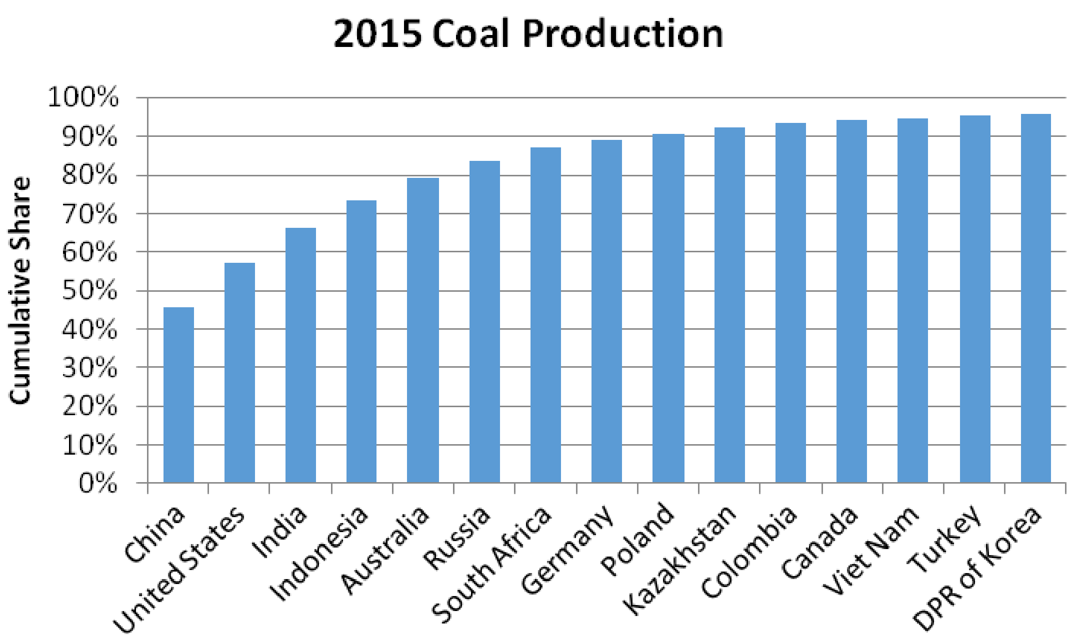

Figure 3: Top Coal Producers in 2015

The coal model will also contain demand nodes for the purpose of determining global trade and determining delivered prices. To the extent that the demand and transformation models are projecting demand at a national or sub-national level, this information could be used to inform the location of the demand nodes, as discussed in more detail in Section 0.

[11] IEA data provided by EIA via spreadsheet "IEA_World coal production.xlsx" February 2017.

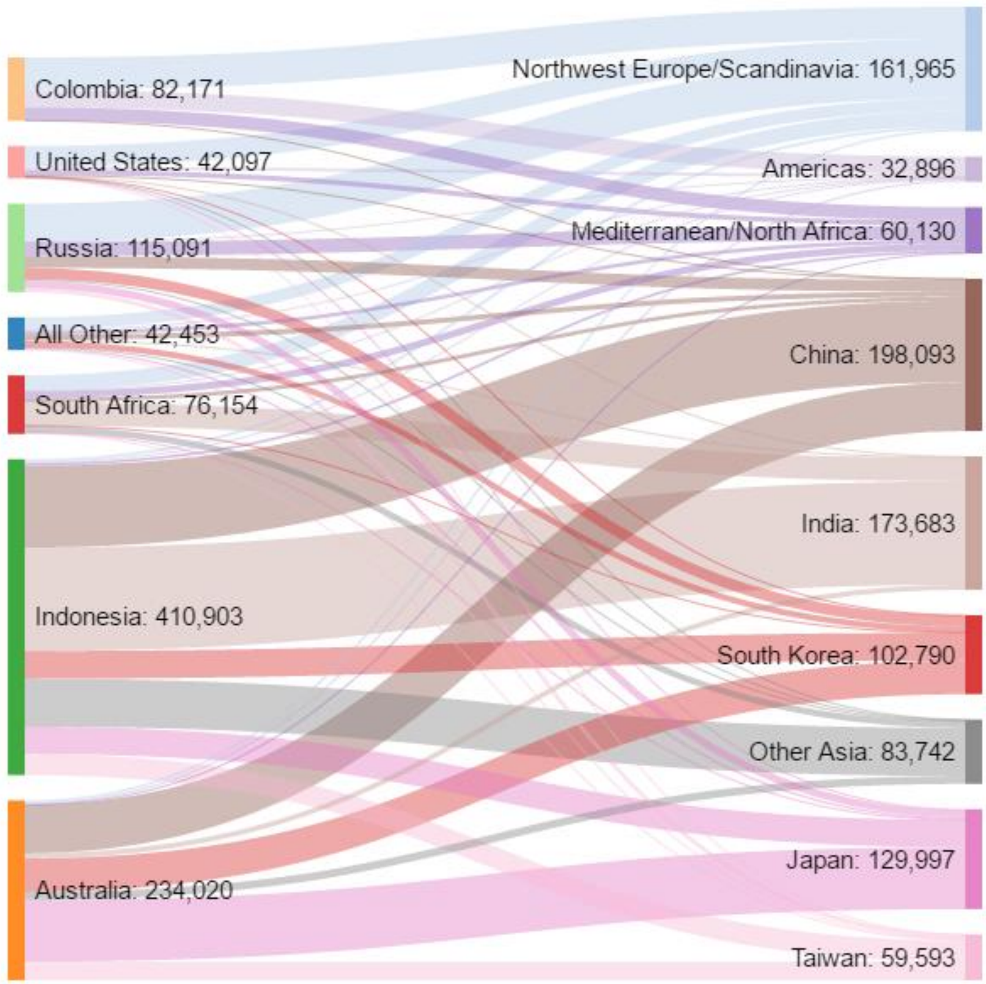

Figure 4: Aggregated Global Trade Flows in 2014, in Thousand Tons per Year

International coal trade is dominated by a relatively small number of exporters and importers as seen in Figure 4.[12] The analysis of global production data with global trade data should give a good starting point for the appropriate level of supply and demand regionality for the International Coal Market Module (ICMM). Furthermore, it may be necessary to further break up certain large producing and/or consuming countries into geographic sub-regions in order to properly represent the market structure. For example, China has large coal deposits in the northwest province of Xinjiang, but the demand for coal is in the east and southeast (Aden, Fridley, & Zhang, 2009). This implies a market choice for China to build infrastructure to move

[12] Created from IHS Energy Global Steam Coal Advisory Service data provided by EIA in the spreadsheet “GSC Steam Coal Trade Matrices.xlsx”..”

coal, build power plants in Xinjiang and the infrastructure to move it, or continue to import coal from the world market. Capturing in the ICMM the cost elements required for this choice would require several supply and demand sub-regions to be modeled for China.

3.2 Representation of Coal Production

Coal production supply curves will be developed econometrically at the greatest appropriate regional detail for each coal type available at that site (usually one type per site). These supply curves represent the current "base" production capacity, which may adjust slightly within a given year without committing capital to expansion. The sites will also be assigned coal qualities, including heat content. Some of the coal quality fields may be placeholders initially if insufficient data are available and are not needed by the rest of the demand and transformation models, for example mercury and sulfur content. These curves will relate coal production levels given existing production capacity to costs or prices. While Hellerworx found that productivity was the most explanatory variable for the two regions in Australia, a full set of variables should be tested in developing curves for other major coal producing areas. A simpler analysis can be conducted for relatively small producing areas.

Production capacity expansion for a given basin may be characterized by an investment cost in dollars per ton of annual capacity addition. The yearly capacity expansion may be capped as well as the total lifetime production of a given basin. This representation has the advantage of being able to separate long-run marginal costs (those including capital costs) and short-term marginal costs (which include fixed and variable costs only). Any additional production capacity selected would then be added to the base production curves for future years. If basin-specific capacity expansion costs are not available, a regional "average" guess may be applied at first and then refined over time if data become available.

Coal production capacity at a given basin could also be reduced if it is consistently underutilized. This would likely be a post-processing step after each solution year, and the thresholds for levels of underutilization and duration of underutilization that would qualify for the retirement of productive capacity would be set by the analyst.

The location of each of these production centers will also be recorded for use by the inland transportation cost algorithm, which is discussed in Section 3.4.1.

3.3 Coal Quality Specification

Each coal deposit around the globe has a unique set of properties which depend on the age of the deposit and the original organic material from which the coal is formed. As the coal deposit ages the moisture content tends to decrease and the carbon content increases. Over extremely long periods of time the organic material transitions from peat to lignite, subbituminous,

bituminous, and finally anthracite coal. Each deposit can be categorized into one of these four coal ranks. The heat content of coal varies dramatically by coal rank. The heat content of lignite coal varies between 4,400 Btu/lb and 8,300 Btu/lb; subbituminous varies from 8,500 Btu/lb to 13,000 Btu/lb; bituminous varies between 11,000 Btu/lb and 15,000 Btu/lb; and anthracite varies between 13,000 Btu/lb and 15,000 Btu/lb. The wide variance in heat content has a significant impact on the cost of transport between the site of production and consumption.

Each coal deposit also has a unique quantity of mineral content. The mineral content has a unique blend and concentration of volatile material associated with the coal. The volatile material may be either heavy metals such as mercury, cadmium, arsenic, and lead, or it may be sulfur or chlorine atoms. These atoms may be part of the organic carbon molecules or they may be in separate inorganic molecules.

The coal characteristics of each coal deposit determine where and how the coal can be used. By far, the major use of coal is for electricity production. “Steam coal” is pulverized, injected into a firing chamber, and burned to heat water in a boiler to create steam. The steam is used by a steam turbine to create electricity by spinning a wire coil in a magnetic field. In addition to the steam and electricity, this process produces solid and gaseous waste. The solid waste is referred to as ash, and the percent of ash produced relative to the coal consumed is referred to as the ash content. The ash content is yet another coal characteristic and is an important specification in the design of coal generators. The inability to deal with the quantity of ash would exclude the use of coals with ash contents above some design specification. The ash also contains much of the volatile content of coal burned. A low concentration of the volatile material can make it possible to use the ash in the production of other products, but at higher concentrations this is no longer an option. As concentrations rise the precautions needed for disposal would also increase. The levels at which these transitions occur depend on the relevant regulations at the point of consumption.

The remaining volatile material would be concentrated in the gaseous waste or flue gas stream. Again the relevant regulations at the point of consumption would dictate which coals can be used. However, investment in equipment to remove toxins and/or pollutants can reduce emissions of the volatile material to acceptable levels and allow for larger selection of coal types. Carbon dioxide (CO_2) represents the majority of the gaseous waste. The carbon dioxide results from the oxidation of the carbon atom. Further, the selection of coals used does little to influence the volume of CO_2 per Btu consumed. Currently, the primary options for reduction in CO_2 emissions are to simply shut down the generator, or make the investment in equipment to capture, compress, and transport the CO_2 to locations where the CO_2 can be sequestered in naturally occurring geological formations or used for enhanced oil recovery.

Currently, most coal fired generators use the pulverized technology relevant to the discussion above. The leading successor to this technology is integrated gasification combined cycle (IGCC). This technology converts the coal into a synthetic gas using heat, pressure, and catalysts. The synthetic gas is then used to fuel a combined cycle type plant to produce electricity. The combined cycle plant is similar to the combined cycle plants that use natural gas. However, using the synthetic gas requires modification because synthetic gas has lower heat content and is more corrosive. The advantages to this technology are greater efficiency, more electricity output per Btu of coal input, greater flexibility in preventing airborne toxins and/or pollutants, and easier capture of CO2. This technology would influence the acceptable coal types while conforming to existing environmental regulations.

The second major use of coal is creation of coke to be used in steel production. This type of coal is known as metallurgical coal. This coal must be bituminous coal with low concentrations of volatile materials and low ash content. To make coke, the metallurgical coal is baked at high temperature in the absence of oxygen. Because of the high value of steel and the limited number of acceptable coal deposits, most of the acceptable coals would be exclusively reserved for coke production. There could be cases where metallurgical coal is used to satisfy demands in the steam coal market and this option may be included in the ICMM; however, the model will likely choose not to do so unless there is a glut of this coal produced in an area where it is expensive to bring to market.

The market for anthracite coal is limited. Mostly it is used in metal fabrication and a small amount is used for direct heat.

Coal demand may expand in the future as an input to processes that convert coal to liquid fuels for use in the transport sector. The choice of coal types that can satisfy these demands will depend on the same coal qualities as discussed above, however, coal-to-liquids (CTL) plants would most likely be mine mouth facilities using a local coal feedstock since it is generally more expensive to transport coal than liquid fuels on a Btu basis.

3.3.1 Recommended ICMM Coal Type Specifications

For all current and potential future applications of coal, the demand and transformation sectors will dictate the variety of coal types required to adequately forecast the demand for coal. In addition, it can be expected that these requirements will evolve over time as the demand and transformation sector models in World Energy Projection System Plus (WEPS+) are enhanced. In order for the ICMM to flexibly accommodate the changes in these coal type requirements, *all known coal deposits should be identified*. Attached to each deposit should be all the coal qualities that define this particular deposit. At a minimum, the list of coal qualities should be:

1. Coal Rank

2. Heat Content
3. Ash Content
4. Moisture Content
5. Sulfur Content
6. Mercury Content
7. Carbon Content
8. Indicator that it can be used to produce coke – metallurgical coal

Add to this list information that will allow for estimation of a supply curve. This list would include:

1. Location, country, province, state, and/or numerical indexes (latitude and longitude, or Global Positioning System, GPS)
2. Depth of the seam (Surface / Deep)
3. Estimation of the size of the deposit and/or current production capability

Compiling this information at the elemental level allows the ICMM to combine deposits both spatially and by similar characteristics to satisfy the coal type requirements from the demand and transformation sectors. As the requirements evolve, the ICMM will be prepared to seamlessly evolve as well, because the data structure to hold and use the data will be in place.

It is anticipated that compiling this information for all coal deposits will be difficult. However, over time, as better data become available, the original estimates can be replaced, and the hope is that the better data will track the need for more refined coal types.

3.3.2 Coal Blending at Demand Nodes

The ICMM will have an option to allow the blending of coal of various qualities at demand nodes to achieve overall quality specifications. The user will have the option of allowing the blending of coal at a given demand node or instead forcing all coal delivered to that demand node to meet the quality specifications individually.

This may be particularly relevant to industrial coal demands in the WEPS+ framework because these demands are for a generic steam and metallurgical coal products without regards to coal quality. One way to handle this in the ICMM is to stipulate a coal quality specification for each industrial demand and let the ICMM choose the optimal blend of coals to meet that demand. Another option is to assume a fixed split of coal types for each regional industrial coal demand. In either case, the delivered price would be calculated as the weighted-average delivered cost over all coal types used to satisfy each industrial demand. The expectation is that the IEMM under development will project demands for coal by quality, although the specification of categories is not yet known.

3.4 Transportation and Logistics

The cost of moving coal from the supply source to the location where it is consumed is often a significant portion of the delivered cost of coal. In choosing the most economic source of coal, the coal transport cost and coal production costs are of equal importance. As a result, an accurate representation of the coal transport network is critical to determining a realistic projection of coal delivery and price. The key factors for coal transport are distance, weight, and available transport mode. The primary modes of transport are rail for land transport, barges for river and lake transport, and ships for ocean transport. Truck transport is an option for short distances, generally less than one hundred miles. For ocean transport, larger ships cost less per ton mile but transport through the Panama Canal can significantly reduce the ton miles required. However, travel through the Panama Canal limits ship size.[13]

To determine transportation costs, locations of demand as well as supply are needed. Because coal fired generators are the largest consumers of coal, tabulating each of these plants and their exact location could serve as a basis for creating demand centroids, or, more ideally, these demand regions can be developed in coordination with the International Electricity Market Module (IEMM) demands. The number of existing plants with coal fired generators is relatively small, less than five thousand. Locating where new plants will be built is speculative, but assuming that new plants will be placed near existing sites is not unreasonable. Existing sites are close to electricity load centers, have transmission access, are near coal transport infrastructure, and are often near water sources for cooling needs. All these resources are also important to new generator facilities as well.

By quantity consumed, coal that is used to produce steel is a distant second to steam coal. This coal is referred to as "metallurgical coal." Unlike steam coal, which can use almost any type of coal, metallurgical coal is relatively clean bituminous coal: low sulfur, low phosphorus, and high clumping properties. This coal is baked at high temperature in the absence of air to produce coke. The coke is added to molten iron to produce liquid steel. While the total quantity of coal consumed in this process is far less than the steam coal consumed, the number of steel mills globally likely exceeds the number of plants with coal fired generators. However, if

---

[13] The expansion of the Panama Canal that was completed in 2016 improved the economics of traversing the canal because the New Panamax vessels can carry approximately 130,000 DWT as compared to the old Panamax ships that could carry 70,000 DWT. This will likely result in more penetration of Atlantic basin coal into Asian markets. (Kaptur, 2014)

possible, tabulating these facilities and their locations would make an excellent basis for determining demand centroids for metallurgical coal.[14]

The remaining coal demand is exceedingly small and includes uses in industrial processes such as the production of carbon fibers and includes a small amount burned for direct heat. This coal use should be ignored except to the extent that the demand modules estimate the consumption. If demand estimates are present, then demand centroids should be estimated based on the demand region represented.

Conversion of coal to liquid fuels or synthetic gas is currently a niche market. However, under some scenarios these processes could undergo rapid growth. There are only a few operating plants worldwide, so the existing facilities are of no help in creating demand centroids. Instead, user judgment will be required to locate future demand centroids based on access to oil and/or gas pipelines and the availability of cheap coal.

One last note, lignite coal cannot in general be transported beyond the border of the coal production area. The high moisture content makes it a transport hazard because it can self-ignite when oxidation occurs. It can support mine-mouth facilities but should not enter the transport network. Exceptions could be made to this rule on a case-by-case basis if desired. Also, some low-rank coal can be processed to make it less volatile. One option is to use the coal to make coal briquettes by compressing the coal into hard blocks. Indonesia is an example of a country that exports a large amount of low-rank coal.

3.4.1 Coal Transport Network for ICMM

The ICMM coal transport network will connect coal supply areas to coal demand centroids. Because most of the coal consumed globally is both produced and consumed domestically (i.e., self-supplied), the coal supply areas and demand centroids may be located within the same country/region (intraregional) or within different countries/regions (interregional). Each coal supply area should be characterized by the available transport modes and by the quality of coal produced. Demand centroids will also be characterized by the mode of transport available. Intermediate nodes are also needed to transfer coal between transport modes. This would create a transport network that connects all coal supply sources to all demand centroids that are accessible through some combination of rail, barge, and port facilities. Ocean transport connects all ports. Finally, imports can move from ports to demand centroids via rail and/or barge. This point-to-point structure has the advantage that existing flows can be explicitly

---

[14] A demand "centroid" is defined here as the weighted-average latitude-longitude coordinates of a given aggregate regional/sectoral demand. The weights used for the average will likely be tons, but could also be Btu.

specified in the matrix and economically encouraged by lower transport rates. This transportation network approach is sufficiently generalized so that it may handle both intraregional and interregional coal transport.

The network representation may be simplified somewhat by creating a transshipment network by introducing intermediate nodes and limiting the number and distance traveled of the transport vectors out of each supply source and into each demand centroid. Long distance transport occurs across a series of links in the transport network. The advantage of this approach is that fewer transport vectors will be required which would reduce the size of the optimization problem thereby reducing solution times.

This formulation implicitly assumes that the transport infrastructure is sufficient to handle the expected flows or that the transport rates are sufficient to allow for investment in additional infrastructure. To capacitate the infrastructure requires two concepts. First, limits on the existing infrastructure must be added to the matrix and new vectors added that allow for expansion. These constraints can be placed on individual nodes to represent limits on the on-load / off-load capabilities at ocean or barge ports. Limiting flows due to congestion along rail lines would be constraints on one or more transport links. Limiting ships, barges, or rail cars would put limits across flows by transport mode and by geographical location. Second, investment decisions need to consider intertemporal issues. The basic transport structure would need to expand to multiple periods so that the investments are consistent with expected changes in demands and therefore make sound economic sense. Investments to match a temporary surge in demand are not likely to be economically sound.

Transportation costs for existing transportation network capacity will be expressed as a US$/unit basis. These costs are origin-destination specific so that costs may vary even within a specific transportation mode. There are several options for the pricing of both inland and interregional transportation. One straightforward way is to create linear functions for transportation costs that depend on variables such as fuel costs and distance traveled. The Hellerworx model described in Section 1.4.2 contains one possible approach for creating international transport cost equations.

For transportation capacity expansion, each option will have an associated capital cost, fixed Operations and Maintenance (O&M) cost, and set of financing parameters that may vary across regions and transportation modes as appropriate. These costs are in addition to the US$/unit variable cost for each transport mode.

Similar to what was discussed in the representation of coal production capacity expansion in Section 3.2, if the explicit capital expansion costs for a transportation link are not available, then the capital expansion may be handled implicitly through a simple US$/ton "levelized" transportation cost subject to a total capacity restriction.

The solution to this problem will yield a price-quantity pair for each demand centroid and for each coal quality specified. This allows, if needed, the ability to create average marginal price and total quantity for each region and coal type specified by the demand and transformation modules. This implies that the ICMM either matches the regionality and coal types in the other modules or it further disaggregates to sub-regions and/or to additional coal types in order to improve the estimate of delivered coal price.

3.4.2 Nodal Geographic Representation

As introduced above, in the analysis and optimization of routing and examining flows of commodities, we find it useful to represent the critical entities in that flow as nodes in a network, or graph. A node is simply an entity that has attributes and can be connected to other nodes by way of links. In flow networks, these links are directed, meaning flows only travel in one direction at a time across the link. These node and link representations have important mathematical characteristics and have been studied and formalized since the 1700s. Modern computer science and analytics make use of this past work to analyze and optimize systems. The result of all this background work is that today we can leverage these systems to perform optimization and organize data about large and complex flow systems using these graph constructs.

In the case of the coal flow network, nodes correspond to physical parts of the energy infrastructure, such as coal production, tankers, ports, rail terminals, and demand centroids. Figure 5 illustrates how a geographic information system (GIS) can catalog the physical infrastructure and process this into a flow network that is suitable for forecasting with iterative optimization models, for instance. These geographic points are linked by rail and vessel routes to complete the transport infrastructure. A key element of the computational framework provided by modern geographic information systems is that it allows us to link sources, transport links, and destinations in ways that are not explicitly collected in survey data on global trade. In this way, the resulting models are greatly improved as they more closely relate to actual ground truth, as well as needing to make fewer assumptions in the model. Rather than relying on rules of thumb or an analyst's tuning of starting parameters for optimization, the needed parameters are calculated directly from the infrastructure geography and characterizations. This geography can be easily represented as a logical graph by connecting the nodes to the relevant links. In practice this requires collecting data through observation or estimation on the routes that the energy commodities take through the supply chain.

3.5 Using a GIS System to Manage ICMM Data

A geographic information system is a system designed to store, update, analyze, and visualize spatial data sets. The ability to view and analyze spatial data in a GIS can reveal relationships, patterns, and data errors that would otherwise be difficult and time consuming to comprehend. There are many input data sets to the ICMM that have a spatial component to them, predominantly those relating to production or transportation infrastructure. The advantage of maintaining these data sets at the finest level of regional aggregation possible is that it enables a modeler to create different regional aggregations for the ICMM by simply choosing which geographical areas fall into each aggregate ICMM region. However, creating and maintaining this fairly detailed spatial data set could be challenging. This is where modern GIS software tools could be very helpful to EIA. A GIS would allow EIA modelers and analysts to:

- Graphically locate coal production and transportation infrastructure
    - Mines
    - Ports
    - Railway terminals
    - Demand centroids
- Graphically add properties to infrastructure
    - Base mine production capability
    - Port capacity
    - Rail link capacity
- Graphically create transportation links that automatically calculate the distance between the origin and destination
- Graphically add taxes and tariffs on the appropriate transportation links
- Graphically create and visualize different regional aggregations of the detailed data and inspect the properties of the aggregate regions
- Make data changes in the GIS software and automatically generate new data sets

While it is certainly possible to create and maintain the ICMM data sets without any GIS software, the ability to visualize all of the data and relationships should make a GIS implementation much more efficient to update the data and verify its correctness. An alternative approach to using GIS software for ICMM data management is found in Section 3.6.

Another advantage to using a GIS system is that updating the data inputs to the model may be done by a subject matter expert rather than a modeling expert. This enables the modeling team to get assistance in keeping the model input data current from those working with the data on a regular basis. The GIS approach also lends itself to distributed upkeep, with each region, or domain expert, responsible for that part of the data set, rather than one entity tasked to update the entire database. The rail infrastructure can be maintained by the transportation sector and the shipping lanes by staff who are familiar with maritime data and updates, and similar aspects.

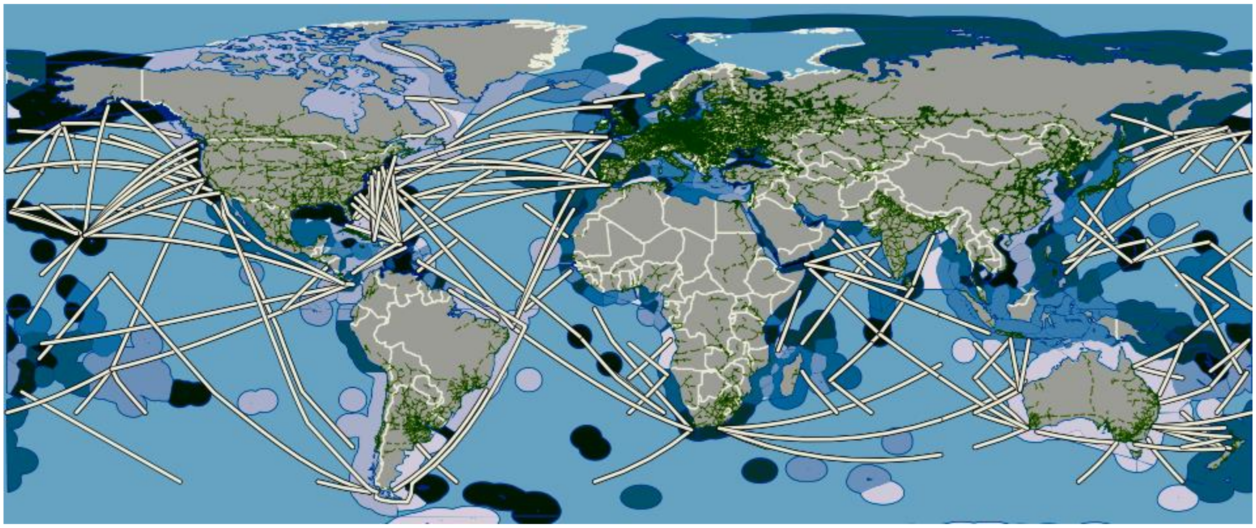

Figure 5: Geographic Representation Facilitates an Accurate Flow Network Representation[15]

It is important to note here that the use of a GIS tool does *not* require that a custom application be created for the ICMM. Free software packages such as QGIS can satisfy most of the needs of the ICMM right off the shelf.

There are currently many sources of data, each representing only a part of the overall picture:

- GIS Data
  - Transport infrastructure
    - Existing capacities
    - Transport costs
  - Points of loading / unloading
  - Points of generation / extraction
  - Points of consumption
- Flow Data
  - Industry survey data
  - Customs and tariff data

[15] Created from import and infrastructure data reported by EIA and IEA, July 2015.

- Forecast Data – outputs from models that forecast demand into the future
- Production Data – volumes of coal entering the market both now and in the future

By leveraging the information held in each of these data sources, we may derive the information necessary for a data-driven model. By generating these models automatically and directly from the source data, the inputs, connections, constraints, and outputs of global energy supply are established. In this way, we define the system automatically, and eliminate a tedious modeling step.

3.5.1 Creating Network Flows from GIS Data

GIS coal-related data sets are maintained by many agencies and provide insight into the locations of global facilities and potential transport links. They do not, however, provide particularly good information on flows and the exact connections between entities. Connections between certain entities are made with the use of outside data sources that confirm a link, such as a survey report stating that 449,313 short tons of coal arrived in Boston, Massachusetts, from Colombia in 2015.[16] This allows us to create a link between the United States and Colombia through this port by conducting some proximity analysis with the GIS to connect the port to the nearby rail network. This approach gives the analyst a method to fill in missing transportation capacity data based on inference from historical data.

Origination and destination nodes, and connections between them, are commonly computed as graph or network databases. Graph databases have some useful properties that facilitate our ability to create inputs to the ICMM optimization model on-the-fly. Graphs consist of nodes and edges. Each can have attributes associated, such as capacities of flow, coal types, and dates of operation (see Figure 6). This type of specificity allows for the creation of a very robust and detailed transportation model without requiring the system to necessarily contain the entire data set of capacities and exact locations.

The resulting graph data set contains potentially hundreds or thousands of nodes and edges, yet can be queried quickly because there are no expensive (time-consuming processing) join operations to conduct, as there would be in a relational data structure.

[16] (US Energy Information Administration, 2016)

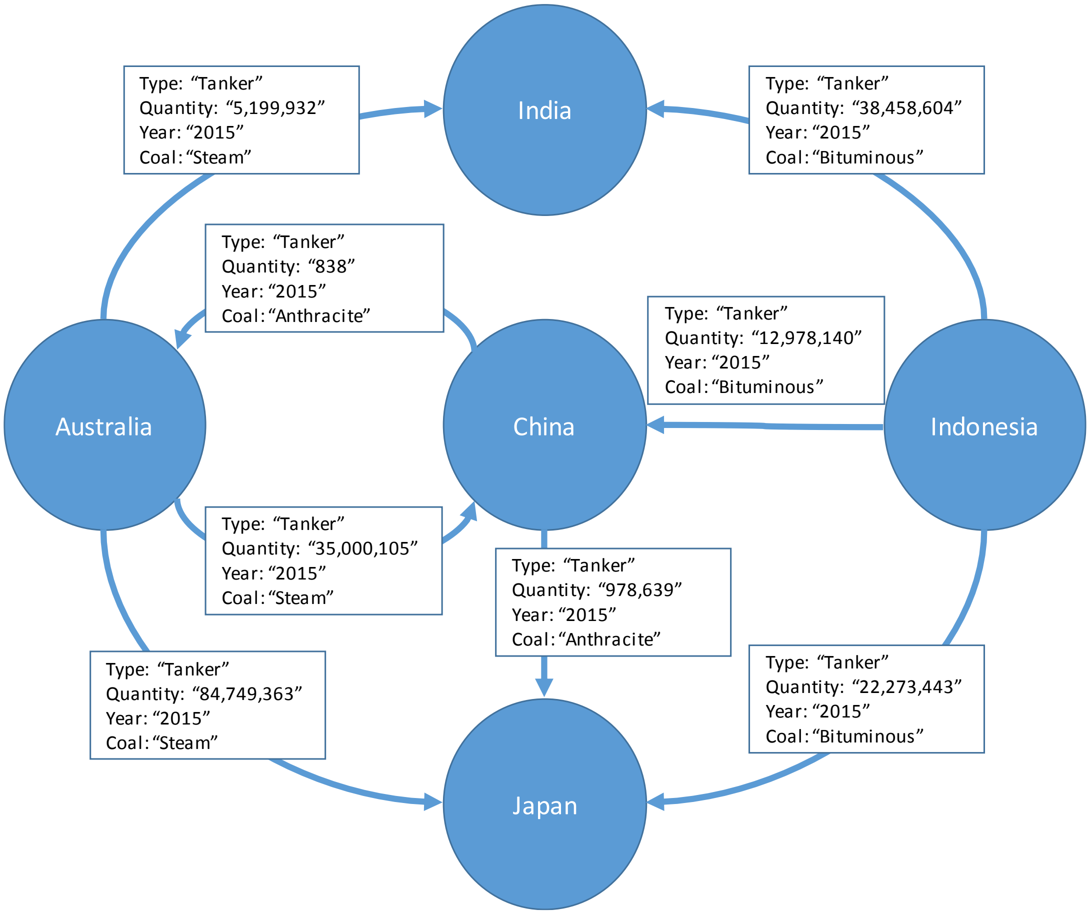

Figure 6: Notional Coal Trade Graph Structure Illustrating Nodes, Edges, and Attributes[17]

3.5.2 Level of Detail

In many cases, exact observation and recording of routes is not feasible and we instead make use of aggregate regional reporting. For instance, we may know the flow of products through a country or multi-country region going from producer to consumer, but not the exact nodes involved. In this case, the modeling can be done just as well at the aggregate level, or the disaggregated values can be estimated through an additional modeling step. It is critical that the model be able to aggregate high-resolution data to accommodate the lowest-resolution data in the system. For instance, coal production can be analyzed at country level, or at an aggregate level such as trade regions. This ability to aggregate easily, even when disaggregated data are only partially reported, is what makes this data-driven approach so flexible.

[17] From EIA spreadsheet 'Exports Coal and coke by country and type.xlsx' provided February 2017.

As long as each entity in the network is geographically located in the GIS, we can apply any regional aggregation desired without re-collecting or estimating any data. The values simply *roll up* to the higher level of aggregation, summing all the constituent parts. One of the challenges of creating a system that has flexible regionality is the automatic calculation of the transportation links that connect the aggregate regions based on the detailed transportation links in the disaggregated data. The GIS software has features built in which handle this aggregation process easily.

Figure 7 illustrates how disaggregated data can be aggregated to reduce the dimensionality of data and focus on the major players in the market. But with no further work, we can show the same data using alternative aggregations. In each case, all we need is to have a definition of which disaggregated regions comprise each aggregated region. The network flow then can populate an optimization model at the required resolution and can run immediately with no further editing needed by a user.

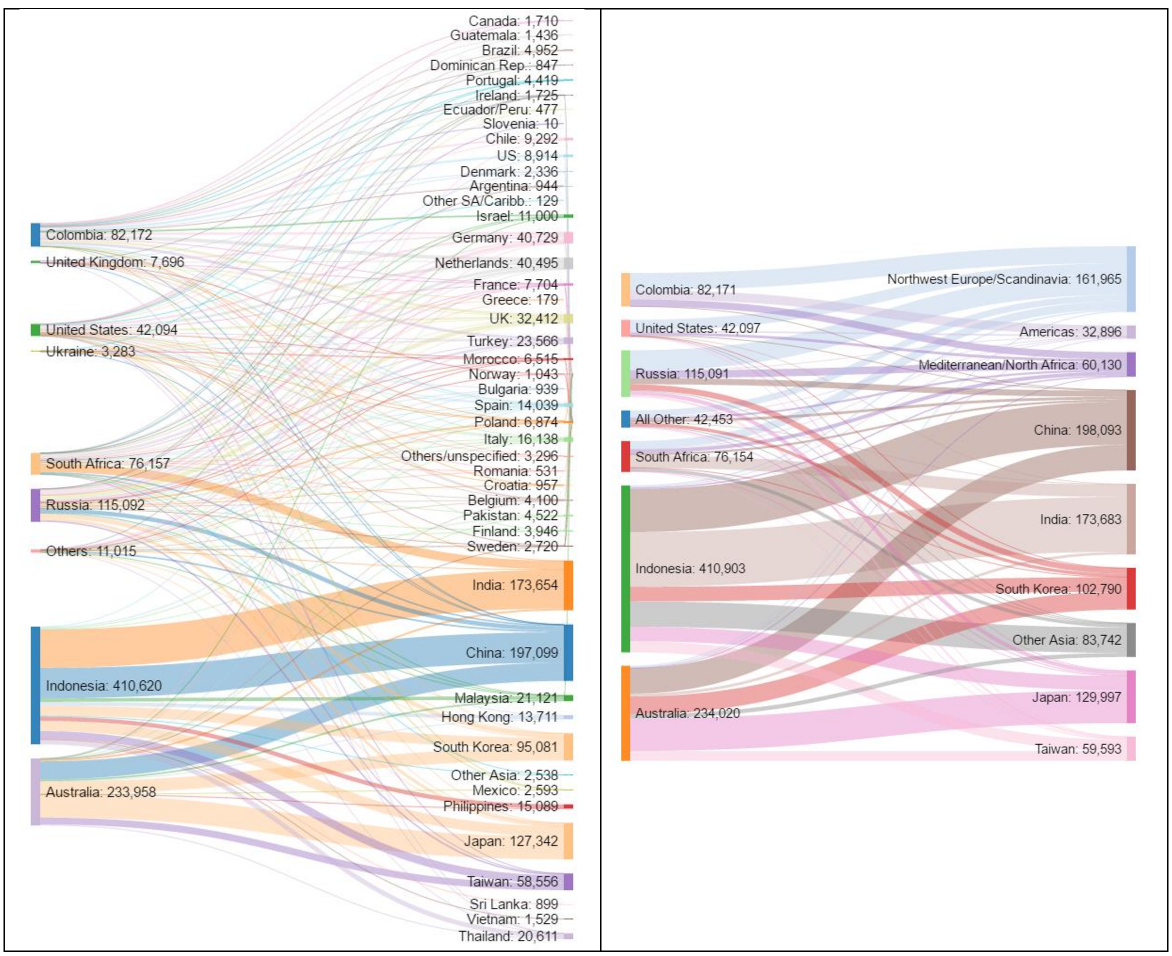

Figure 7: Disaggregated vs. Aggregated Steam Coal Trade Flows

There are two primary advantages of modeling at the finest level of geospatial detail. The first is that the graph algorithms benefit from the specificity of the more accurate graph, and then can report results at an aggregate level. Calculating at the country level and aggregating to the WEPS+ regional level yields better results than calculating at the WEPS+ regional level from the start. The second advantage is that the more disaggregated model can be *rolled up* or aggregated to whatever level is appropriate for the question being posed of the model - national level aggregation for trade purposes, and state level for reporting purposes, for example. The initial graph optimization model does not need to be altered or re-run to accommodate these various on-the-fly reporting aggregations.

3.6 Managing ICMM Data Without GIS Software

The data management functions of the ICMM may be adequately handled without any GIS software tools. We may use a combination of text files (CSV), database software (SQLite), and Python to create, store, modify, aggregate the input data, and report results for the ICMM. The raw disaggregated spatial data could be entered into CSV data files manually and then imported into a SQLite database. The SQLite database enables the SQL query functionality that, together with Python, will be needed to aggregate the raw detailed data into the desired regional aggregation. In addition to serving as the interface to the SQLite database and performing the data aggregation, Python could also be used to:

- Calculate the distance represented by each of the transport links (assuming the latitudes and longitudes of each origin-destination pair are available)[18]
- Perform other necessary data preparation and manipulation
- Create automatic validation checks on the input data
- Create the data input files for the ICMM LP model
- Generate reporting tables and graphs

In addition, you could use Python and the Mapnik library to create map displays from the input and solution data for verification and display purposes.

---

[18] This is needed since the transport costs will likely be functions of distance traveled.

4 Solution Methodology

4.1 Modeling Approach

The International Coal Market Module (ICMM) will be a generalized network flow optimization problem that seeks to satisfy all coal demands at a minimum cost subject to transportation capacity constraints, with the option of building additional transportation capacity as needed and if economically viable. The modeling framework would use a linear programming (LP) approach initially, but could be switched to a mixed-integer programming (MIP) solution approach if certain types of capacity expansion decisions need to be modeled as all-or-nothing decisions.

The ultimate size of the problem solved depends on the level of aggregation. This includes the number of coal types, the number of supply locations, the number of demand centroids, and the number of transport modes. If the inputs have been created and maintained at the lowest level possible, the work to aggregate to a level that complements the requirements of the demand and transformation sectors should be minimal.

Key advantages to using an LP approach for the ICMM model include:

- Linear programming is an extremely well-understood modeling framework that has been studied and utilized extensively for decades. As such, there is a large knowledge base of support, both within and outside EIA, for this modeling approach.
- LPs that are well-posed are virtually guaranteed to converge to an optimal solution. Other approaches can fail either to converge completely or find varying local optima starting from different initial solutions. This becomes important when you have several groups of people using a model for scenario analysis whose results must be comparable and repeatable.
- LP solution times tend to be among the fastest across optimization models. Models with millions of variables and constraints can be solved in short order on most laptops nowadays, especially in instances which have good starting solutions.
- Used in conjunction with the proposed front-end geographic information system (GIS), the data handling duties are dramatically reduced and the robustness of the model specification system improved.

Disadvantages to using an LP approach:

- Linear programs can potentially yield “knife-edge” solutions where small changes in coefficient values can lead to large changes in solution values.

- Linear programs use a single global objective function in which it is assumed that all market participants are behaving in a centrally coordinated way that maximizes global benefit (or minimized global cost.)
- Linear programs cannot represent imperfect competition in markets where a few producers dominate the market and can exercise market power.

In competitive markets, the marginal values from the LP results can be used to set prices. In markets with imperfect competition, some additional modeling structure may be required as discussed in Section 4.4 below. Further, the investment costs input to the model can be adjusted with alternative localized discount rates to accommodate the variations in costs-of-money based on assumptions as to whether the region is assumed to be free market driven, sovereign or country subsidized, or other. This flexibility in establishing pricing regimes in the model is key to its being able to provide the appropriate price signals to the demand models and to motivate the expansion of the network based on regional market factors.

4.1.1 Endogenous Coal Production Expansion and Transport Capacity Additions

The projection time frame of WEPS+, currently to year 2050, requires the modeling framework be capable of addressing the economically efficient expansion of the coal production and logistics infrastructure. Large investments in the development of coal basins, port handling facilities, transport vessels, and other infrastructures require a view of the future capacity and capacity utilization needs. The ICMM will support endogenous production and transportation capacity additions for the various transportation modes in order to properly represent scenarios in which global demands for hydrocarbon transportation are changing.

Economically efficient expansion addresses the least-cost short-term and long-term utilization of current and expanded capacity, recognizing the existence of historical out-of-market contracts/deals. That is to say, transportation capacity expansion and rationalization will be determined by the least-cost mix of all costs, including capital, Operations and Maintenance, and fuel needed to satisfy a particular transportation demand slate over the projection horizon in a given model year, subject to any historical overrides (for example, forced use of selected elements of the transportation system due to historical contracts).

Constraints on the rate of production and transportation capacity expansions will be introduced where appropriate. For example, Indian coal imports are in part driven by the inability to transport domestic coal supplies to domestic power generators.

Differing approaches to capacity expansion may be used across different assets. A continuous LP approach will most likely be sufficient for most capacity expansion decisions. For example, fractional tanker ship builds should not be problematic in a long-term forecasting model such as the ICMM because each tanker has a relatively small transport capacity. However, we may want the ability to make integer choices when it comes to import/export facility capacity expansion due to the large capital cost requirements. It is very easy in modern modeling languages such as the Advanced Interactive Multidimensional Modeling System (AIMMS) to switch back and forth between variables being defined as continuous or integer-valued, and so experimenting to gauge the trade-offs between model run time and solution quality is a low-cost exercise.

4.1.2 Temporal Representation

The ICMM will be an annual forecasting model with the planning horizon synchronized to the rest of the WEPS+ because it will depend on other models for inputs. However, the ICMM itself is data-driven with respect to time intervals, and so, one would only need to change the input data sets to use a different set of time intervals in the model; no code changes should be required.

Because of the foresight requirements of the capacity build decisions, data for future years past the last year of the WEPS+ forecasting horizon also need to be specified or estimated via a specifically adopted approach. One option is to trend out values past the forecast horizon using an assumed or calculated growth rate (i.e., an average growth rate calculated from some number of years prior to the last forecasting year). Another option is to simply assume the value in the last forecasting year persists indefinitely.

4.2 Pricing Approach

The ICMM must produce prices at several points along the coal supply chain (Figure 8):

1. Mine-mouth prices at each coal supply node
2. Free-on-board (FOB) export prices at each export terminal or port
3. Cost, Insurance, and Freight (CIF) import prices at each import terminal or port
4. Delivered prices by sector in each demand region

Delivered prices will be reported in $/MMBtu for use by the other WEPS+ models. The intermediate prices could be reported in either tonnage or energy units. Taxes and tariffs will be applied on appropriate transport links along each step of the supply chain. Regions that do not import or export any coal will have only mine mouth prices and delivered prices by sector. These self-supplying regions will still have a full representation of coal production and demands in the ICMM, and the delivered prices will the costs of production plus inland transportation costs.

Prices will be projected by the ICMM supply and demand sub-regions and then aggregated to the 16 WEPS+ regions.

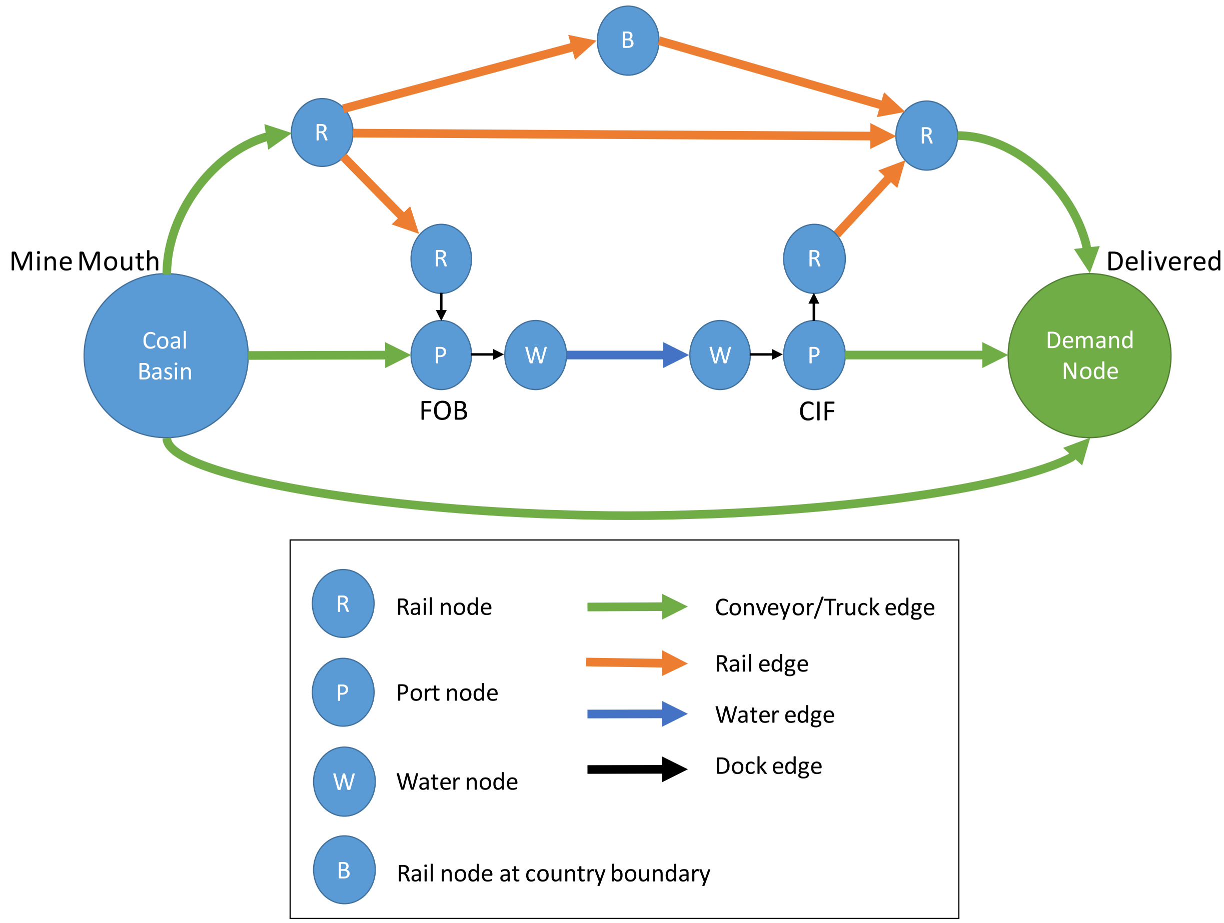

Figure 8: Price Points Along the Coal Supply Chain

4.2.1 Marginal Cost Based Pricing

Marginal cost pricing reflects those markets where producers bid into a competitive market and all producers are paid the clearing price that matches supply to demand. It is the assumption in this Component Design Report that the global coal market is a competitive market and therefore the marginal prices from the LP model will be used to set prices in general. A brief description of how the model might be able to address alternative market assumptions is included in Section 4.4 below.

4.2.2 Contract Pricing and Tariffs

Known contracts for delivery of coal can be accommodated in the ICMM framework through the setting of bounds in the LP model. A known contract to deliver a volume of a coal between an origin and destination along a specific transportation link may be dealt with by setting lower bounds on the transport of coal on that link and thereby taking the decision away from the ICMM LP model. Note that one of the advantages of using a constrained optimization approach is that it can signal the economic merit (above or below cost) of existing contracts on the transportation network. Further, note that existing tariffs or fees can be placed on the network to represent historical contract arrangements. The solution obtained with these "costs" can then be compared to that of an economic cost-driven solution to establish the "value of the contract." This can inform both how long the contract should be assumed to persist and how the network will respond as the forced costs or flows are taken off the network.

A related set of constraints that may also be added to the ICMM would be upper limits on imports or exports of coal. An example of this would be a country setting a limit on coal imports to protect its domestic coal producers. Of course, this would most likely be at the expense of its coal consumers because the model would only choose to import coal if it were cheaper than the delivered cost of domestic supplies.

4.3 Foresight and Investment Planning

As introduced above, the ICMM will employ a multi-period expansion planning approach that simultaneously optimizes the short-term utilization of existing coal production capacity and/or transportation capacity to meet current global coal demands, as well as the long-term capital budgeting schedule for coal production expansion to meet future coal demands. The dimensions of the investment decision that are most important include:

- Forward-looking with respect to expectations of spatially specific quantities (entry and exit point), constraints on utilization, costs, and technology (e.g., size of tankers, Panamax constraints)
- Ability to handle different risk profiles for each region and type of asset (e.g., mines, ports)
- Ability to incorporate different capital asset lifetimes
- Ability to address differing market conditions (i.e., fully competitive versus sovereign state)

The existing Liquid Fuels Market Module (LFMM) and Electricity Market Module (EMM) in the National Energy Modeling System (NEMS) each incorporate a multi-period planning model framework to successfully address each of these dimensions.

It is critical that foresight is included into a forecasting model that has capacity expansion decisions. Large capital investments with long construction lead times, such as the development of a new coal basin or major port facility, are done with an expectation as to the size and direction of the future markets. An example from natural gas markets would be to determine whether or not to build a natural gas liquefaction facility to export liquefied natural gas (LNG) from the United States to Europe today if one expects a new pipeline from Russia to

Europe to come online in the near future that would undercut one's prices and strand one's asset. Specifically, the advantages of foresight include:

- Allows for making investment decisions with some expected knowledge of future market changes or policy shifts that may impact investment choices
- Reduces the chance of building large assets that are then stranded or underutilized due to future fluctuations in prices and demands
- Once current assets are rationalized, "smooth" capacity expansion trajectories will result without major year-to-year changes in coal production

Designing and implementing the ICMM as a multi-period model gives the analyst the ability to experiment with varying planning horizons and their impact on decisions/projections. Understanding that making significant investments involves incorporating the opportunity costs of money into the decision framework is important. By convention, we apply an adjusted discounted net present value to compare the competing investment options. To construct this view of the future, expected values of the key inputs are taken from the prior solution iterations of WEPS+ with some framework for extrapolation beyond the planning horizon. Note that alternative constructs of expectations could be implemented without significant adjustments to the modeling framework.

Because WEPS+ solves each of the sector models individually through the projection period and then iterates among them, coal demands and fuel prices from prior solutions of the demand and fuel supply models can be used to inform the ICMM. Although foresight will be used, the LP will solve each forecast year, as in the EMM, rather than optimize across all future periods at one time. This allows the model to more reasonably reflect investor behavior in the light of expected future economic and regulatory conditions. The annual approach would also allow the models of WEPS+ to communicate on an annual basis rather than once per cycle if that proved to be a better solution in the future for convergence.

The proposed ICMM LP will use multiple time periods, much like the LFMM and Electricity Capacity Planning (ECP) module of the EMM in NEMS. As in the ECP, we propose that the ICMM LP have three planning periods:

- The current year to capture production from existing mining capacity
- The next year in which capacity build and retire decisions will be made
- A third period representing the remaining years in the planning horizon that informs the model regarding evolving economics and policies over the long term and that allows for the model to delay expansion or retirement decisions

The number of periods will be specified in a flexible manner so that the user can experiment with the number of planning periods as well as the number of years in each planning period.

The LP framework is designed to return annual solutions that:

- Meet specified coal demands by coal type in each region and time period at a minimum cost
- Satisfy transportation logistics capacity constraints
- Determine the lowest price that can satisfy all coal demands while meeting investors' minimum rate of return investments requirements

Because this is a multi-period model, all prices/costs in the LP formulation are calculated as the net present value (NPV) nominal unit price discounted to the beginning of the full projection horizon. This value is calculated as the NPV of the values associated with each year of that planning horizon and then discounted to the first year of the projection horizon.[19] The first-period costs will simply be the costs from the current year in nominal dollars. Details regarding the use of NPV calculations in the ICMM may be found in the Appendix.

For example, the following steps would be taken to calculate an NPV US$/ton/day coal production expansion cost in a given planning period:

1. A US$/ton/day nominal annuity will be calculated using an approach analogous to the one used in the EMM or LFMM.
2. The NPV of the stream of annuities is calculated over the number of years from the beginning of the current planning period to the end of the investment horizon using a weighted-average cost of capital (WACC) as the discounting rate.
3. This NPV value is then discounted back to the first-period nominal dollars.

In addition, all constraint right-hand sides and bounds (e.g., coal demands, transportation constraints) in a given planning period are calculated as an NPV average value. This average is computed as the NPV of the value in each year associated with a given projection period, divided by the NPV of one unit of product over the same time period. This is equivalent to taking the NPV of unitized values (i.e., the NPV of the values divided by the NPV of the units.)

An alternative approach to capacity expansion may be considered. The ability to explicitly model coal production capacity expansion at a fine level of aggregation depends on the

[19] To account for the time value of money, the discounting of the impacts of future changes in the cost of key inputs and constraints needs to be included in the model formulation. The standard approach to modeling this economic concept is through the use of the NPV calculation.

availability of the financial data relating to new mining investment in each coal-producing basin. Absent such data, the additional computational overhead imposed by the multi-period planning may not pay off in terms of additional modeling fidelity. If the capital expansion costs for a given basin are not available and reasonable estimates cannot be created, then the capital expansion may be handled through additional steps on the base production supply curve using an estimated "levelized" cost. If these capacity expansion supply curve steps are used in a given basin, then the base production curve for that basin would be shifted for future years to account for the expanded production capacity.

4.4 Modeling Individual Producer Strategy and Imperfect Competition

While linear programming is an important and powerful tool for analyzing economic problems, there are cases where it is not appropriate. One such case is modeling a market in which a few market players are dominant. In this case, market participants may have some degree of market power and so are not pure price-takers.

Under perfect competition represented in an LP model, the transfer prices of energy commodities between each supply/transformation sector would be set according to the dual variables on the commodity transfer balance rows which would reflect the true costs of production. We have assumed to this point that the global market for coal is perfectly competitive (or close enough to model it that way), however the possibility exists that some of the major players in the global coal market may exhibit market power or other non-competitive behaviors. Market power means that some producers would be able to set marginal prices equal to marginal revenue instead of marginal cost. This situation is not properly represented by a linear programming framework. In this case, each producer would have its own profit-maximizing objective function, and these cannot be combined to form a single global objective function in an optimization model.

An example of this would be a large producer and consumer, such as China, instituting a policy of restricting export supply to the global market to keep domestic prices lower in order to maximize the sum of producer and consumer surplus and export revenues.

The decision of a producer (or subset of producers) to set production levels in order to maximize their own revenue or influence the global market prices to maximize personal welfare entails modifying the dual variables on the production balance rows in such a way that the marginal transfer prices do not equal the marginal production costs. Optimizing over primal and dual variables concurrently requires a mixed-complementarity problem (MCP) formulation. See (Murphy, Pierru, & Smeers, 2015) and (Paulus, 2012) for greater detail on the use of MCPs to model energy markets with imperfect competition.

Mixed Complementarity Problems (MCP): An Overview

Complementarity problems are used to solve market equilibrium problems with multiple profit-maximizing players. In the traditional model of imperfect competition (Cournot, 1938), firms decide independently what quantity to produce to maximize their own profit given the output of other firms. When no firm can unilaterally improve their own situation, a so-called Nash equilibrium is reached. It has been shown that any optimization problem can be represented by an equivalent mixed complementarity problem (MCP) (Haftendon, 2012). The MCP of a linear program comprises the constraints from both the primal and dual linear programs, and the "complementary slackness" conditions which state that:

- A dual variable can be positive only if the associated primal constraint is binding
- A primal variable can be positive only if the corresponding dual constraint is binding

Furthermore, MCPs are more general than optimization problems because they may include side constraints which allow the representation of market behavior that is not purely competitive and contain explicit statements of price relationships and complementarity conditions where the prices (duals) appear directly as variables.

Linear programming is still the preferred tool of choice when the appropriate market conditions are met for several reasons:

- Setting up an LP is faster because you do not have to write down the associated dual program
- Solution algorithms for solving LPs are much faster than the ones for solving MCP problems, and they can handle much larger problem sizes.

It is important to note here that the effort to formulate the global coal market as a perfectly competitive LP would not be wasted if it were later decided to model non-competitive behavior. In fact, the first step to creating an MCP formulation is to model each player as an individual LP and then combining and supplementing these with the complementarity equations. It is therefore a straightforward exercise to convert the LP model to an MCP model if desired.

The general approach may be summarized as:

1. Develop each regional primal LP model.
2. Store the dual LP problem for each region.
3. Combine the primal and dual equations of all regions, along with their associated complementarity conditions, into a single MCP.
4. Modify the appropriate price/quantity relationships and complementarity conditions to represent the non-competitive conditions such as market power or administered transfer prices.

An alternative to the MCP approach would be to iteratively solve each regional LP separately and repeatedly until convergence is reached. This solution method is a Gauss-Seidel algorithm and is currently the solution approach of the NEMS model. While each iteration “cycle” of solving all regional LPs would likely be faster than solving the combined MCP problem, the potentially large number of iteration cycles required for convergence of the iterative approach may extend the total solution time.

4.5 Sensitivity Scenarios and Model Response

The ICMM should be able to be responsive to a wide range of future coal demands. These demands are currently in flux globally. Environmental policies, as well as low-cost natural gas and renewable energy have led to reductions in coal demand in many Organization for Economic Cooperation and Development (OECD) countries, while increasing economic growth and electricity generation have increased demand in many developing countries, especially India, China, and the rest of Southeast Asia.

Future demands for coal are also dependent on the economic viability of carbon capture and storage (CCS) in conjunction with climate change mitigation policies. Without viable CCS, coal demand would need to decline if countries adopt stringent greenhouse gas (GHG) emission reduction targets, while successful CCS deployment could allow for an expansion of coal use depending on the policy stringency. Therefore, the ICMM should be able to respond to either coal demand growth or decline overall and by region.

The *International Energy Outlook*’s (IEO’s) traditional alternative scenarios of high/low economic growth and oil prices will also lead to changes in coal demands. Oil prices should both directly impact coal prices because oil is a significant cost in coal mining and transport and indirectly through inter-fuel competition between the fuels. The ICMM needs to project appropriate coal prices for these alternative levels of demand so that the demand and transformation models can project the relative attractiveness of using and investing in coal technologies versus other options. In high oil price cases, a particular wild card is the economic viability of coal-to-liquids production that could have significant implications for coal demands and need for production and transport capacity expansion.

4.6 Mathematical Specification of the LP Model

4.6.1 Index Definitions

$d \equiv$ *Index of coal demand nodes*
$\varepsilon \equiv$ *Index of coal quality properties*
$m \equiv$ *Index of transportation modes*
$n \equiv$ *Index of coal supply nodes*
$t \equiv$ *Index of planning periods*
$x \equiv$ *Index of coal production supply steps*

4.6.2 Column Definitions

These are the decision variables.

1. $P_{n,x,t} \equiv$ Coal production at supply node n on supply step x in planning period t in M tons.
2. $T_{i,j,n,\mathrm{m},t} \equiv$ Transport between nodes i and j of coal produced at n on transport mode m in period t in M tons.
3. $D_{d,t} \equiv$ Demand at node d in period t in MMBTU.
4. $B_{d,\varepsilon} \equiv$ Blended coal quality requirement ε at demand node d.
5. $S_{d,n,t} \equiv$ Supply of coal produced at node n in demand node d in period t.
6. $M_{i,j,m,t} \equiv$ Existing Transport capacity between i, j on mode m in period t.
7. $TB_{i,j,m,t} \equiv$ Transport capacity builds between i, j on mode m in period t.
8. $V_{m,t} \equiv$ Total marine vessel capacity of vessel type m in period t.
9. $VB_{m,t} \equiv$ Marine vessel capacity builds of vessel type m in period t.
10. $TC_t \equiv$ Total transportation costs in period t.
11. $VC_t \equiv$ Total coal production variable costs
12. $CC_t \equiv$ Total coal production capital costs
13. $PB_{n,t} \equiv$ Production capacity builds at node n in period t.

4.6.3 Objective Function

The objective function seeks to minimize the net present value of total costs over all planning periods (t) in nominal dollars in order to satisfy all regional demands for coal. This yields the lowest price that will support (cover the costs) all the current and future capacity and provides the required minimum rate of return on all capital investments. The total costs include capital expansion costs, variable production costs, fixed production costs, and transportation costs, as well any applicable fees (taxes and tariffs). In the first period, available production capacity is restricted to the base production curves, and the cost minimization determines the lowest cost operation of that capacity. This is the lowest price necessary to cover the cost of the marginal producer.

$$\text{Minimize} \sum_t \{TC_t + VC_t + CC_t\}$$

4.6.4 Description of Constraints

Transportation costs. For each planning period, the total transportation cost equals the sum of transportation costs across all transportation modes. Transportation costs are a function of fuel prices and distance traveled. Fuel prices reflect current year prices as well as projected future values. Fuel consumption is characterized by transportation mode as well as fuel type.

$$TC_t - \sum_i \sum_j \sum_n \sum_m \{T_{i,j,n,m,t} * \tau(i,j,m,t) + M_{i,j,m,t}\}$$
$$= 0 TC_t - \sum_i \sum_j \sum_n \sum_m \{T_{i,j,n,m,t} * \tau(i,j,m,t) + W_{i,j,m}\} = 0$$

The transportation cost function will be a linear function of fuel prices and miles traveled:

$$\tau(i,j,m,t) = D_{i,j,m} * C_{i,j,m} + \alpha_m * FC_{m,t}$$

Where:

$\tau(i,j,m,t) \equiv$ *NPV nominal transportation cost between nodes (i) and (j) on transportation mode (m) in planning period (t) in US$/ton*
$D_{i,j,m} \equiv$ *Shipping distance between nodes (i) and (j) on transportation mode (m)*
$C_{i,j,m} \equiv$ *Base shipping cost between nodes (i) and (j) on transportation mode (m)*
$FC_{m,t} \equiv$ *Fuel cost for transportation mode (m) in period (t)*
$\alpha_{m,t} \equiv$ *Fuel cost adjustment coefficient for transportation mode (m) in period (t)*
$M_{i,j,m}W_{i,j,m} \equiv$ *Total taxes and tariff markups between nodes (i) and (j) on transportation mode (m)*

Variable Costs. The total variable operating cost over all coal production basins.

$$VC_t - \sum_n \sum_x VOM_{n,\text{x},\text{t}} * P_{n,x,t} = 0$$

Where:

$VOM_{n,x,t} \equiv$ *NPV nominal variable production cost in US$/ton for coal production node (n) on supply step (x) in planning period (t)*

Capital and fixed costs. The total fixed cost equals the sum of all capital plus fixed operating costs from capacity expansion over all coal production basins and transportation infrastructure builds. Existing capacity has only fixed operating costs, while investment and fixed operating costs are attributed to new capacity.

$$CC_t - \left\{ \sum_n XPC_{n,t} * PB_{n,t} + \sum_n \sum_x FOM_{n,t} * P_{n,x,t} \right\} - \sum_i \sum_j \sum_m XTC_{i,j,m,t} * TB_{i,j,m,t} - \sum_m XVC_{m,t} * VB_{m,t} = 0$$

Where:

$FOM_{n,t} \equiv$ *NPV of nominal fixed operation and maintenance costs for coal production node (n) in planning period (t)*
$XPC_{n,t} \equiv$ *NPV of nominal annualized capacity expansion capital costs and fixed costs over the remaining years of the planning horizon for coal production node (n) in planning period (t)*
$XTC_{i,j,m,t} \equiv$ *NPV of nominal annualized capacity expansion capital costs over the remaining years of the planning horizon for the transportation link defined by origin node (i), destination node (j), and transport mode (m) in planning period (t)*
$XVC_{m,t} \equiv$ *NPV of nominal annualized capacity expansion capital costs over the remaining years of the planning horizon for the marine vessel mode (m) in planning period (t)*

Minimum Delivered Coal Quality. The coal delivered to demand node (d) must meet the set of all minimum quality requirements (ε_{min}).

$$\sum_n Q_{n,\varepsilon_{min}} S_{d,n,t} - B_{d,\varepsilon_{min}} \sum_n S_{d,n,t} \geq 0; \quad \forall \ d, \varepsilon_{min}, t$$

Where:

$Q_{n,\varepsilon_{min}} \equiv$ *Coal quality (ε_{min}) property value of coal produced at supply node (n)*
$B_{d,\varepsilon_{min}} \equiv$ *Minimum coal quality (ε_{min}) property requirements of coal demanded at demand node (d)*

Maximum Delivered Coal Quality. The coal delivered to demand node (d) must meet the set of all maximum quality requirements (ε_{max}).

$$\sum_n Q_{n,\ \varepsilon_{max}} S_{d,n,t} - B_{d,\varepsilon_{max}} \sum_n S_{d,n,t} \leq \quad 0; \quad \forall \ d, \ \ \varepsilon_{max}, t$$

Where:

$Q_{n,\varepsilon_{max}} \equiv$ *Coal quality (ε_{max}) property value of coal produced at supply node (n)*
$B_{d,\varepsilon_{max}} \equiv$ *Maximum coal quality (ε_{max}) property requirements of coal demanded at demand node (d)*

Aggregate Transported Supply over Modes. The supply of coal produced at supply node (n) in demand node (d) must equal the amount transported in over all transportation modes.

$$S_{d,n,t} - \sum_i \sum_m T_{i,d,n,m,t} \quad = \quad 0; \ \ \forall \ \ d, n, t$$

Transport Flow Balance. The amount of coal produced at supply node (n) entering a given transshipment node (i) over all transport modes must equal the amount exiting over all transport modes for every time period (t).

$$\sum_m \sum_j T_{i,j,n,m,t} - \sum_m \sum_j T_{j,i,n,m,t} = 0; \ \ \forall \ \ i, n, t$$

Transport Capacity Constraints (non-marine). The total quantity of coal transported over transportation link (i,j) on transportation mode (m) must be less than or equal to existing transportation capacity plus capacity expansion builds.

$$\sum_n T_{i,j,n,t,m} \leq \quad M_{i,j,m,t} + TB_{i,j,m,t}; \ \ \forall \ \ i, j, m, t$$

Where:

$M_{i,j,m,t} \equiv$ *Existing Transport capacity between (i, j) on mode m in time period (t).*

Production-Transport Balance. The quantity of coal exiting coal supply node (n) must be greater than or equal to the quantity transported out.

$$PB_{n,t} + \sum_x P_{n,x,t} - \sum_j \sum_m T_{n,j,n,m,t} \geq 0; \ \ \forall \ \ n, t$$

Satisfy Demand in Btu: Coal demands in MMBtu must be satisfied at each demand node.

$$\sum_{x} Q_{n,\varepsilon_{BTU}} S_{d,n,t} \quad = \quad D_{d,t} \ ; \ \ \forall \ \ d,t$$

Where:

$Q_{n,\varepsilon_{BTU}} \equiv$ *Coal energy content in MMBtu/ton of coal produced at supply node n.*

$D_{d,t} \ \ \equiv$ *Demand at node d in period t in MMBtu*

Marine Vessel Capacity Constraints: The total amount of coal shipped on marine mode (m) over all origin-destination pairs must be less than or equal to the total global capacity of that mode.

$$\sum_{i}\sum_{j}\sum_{n} T_{i,j,n,m,t} \leq \quad V_{m,t} + VB_{m,t}; \quad \forall \quad t, m \in Marine \quad Modes$$

Where:

$V_{m,t} \equiv$ *Total marine vessel capacity of vessel of type (m) in period (t)*

5 Implementation Approach

This section on the implementation approach describes the software packages to be used for the various facets of the International Coal Market Module (ICMM) preparation. These packages must work together to complete the following tasks:

1. Accept user input from either a front-end graphical user interface (GUI) or a control file indicating which files and parameters should be used for a given model run[20]
2. Read in data sets and prepare data for use in the ICMM linear programming (LP) model based on user-specified parameters
3. Solve LP
4. Prepare solution results
5. Write solution results to a central database
6. Display results

5.1 Software Considerations

It is a requirement of the ICMM to use Python for any data manipulation and model interfaces and to use the Advanced Interactive Multidimensional Modeling System (AIMMS) for optimization.

5.2 User Interface

While the raw form of many of the ICMM tables will remain as comma-separated values (CSV) files, our intention is to implement a system that (1) brings all of the data that are used by the ICMM (including the restart variables) into an SQLite database and (2) will have a front end that allows the user to select parameters to be used to disaggregate the 16 World Energy Projection System Plus (WEPS+) regions to whatever sub-regions are desired. Once the user is satisfied that the sub-regional setup is acceptable, the program will produce the data inputs needed by the AIMMS LP model (see next section regarding model generation). The model will then solve and provide a standard set of reports as well as an output SQLite database.

There are many possibilities if EIA makes the decision to create a front-end GUI application for the ICMM. For example, the front-end GUI could be used to set up and launch ICMM runs only, or it could also incorporate some geographic information system (GIS) functionality that provides feedback to the user relating to the specified input data set. This input data could be viewed at the WEPS+ regional level, or at the further disaggregated sub-regional level at which

[20] This is analogous to the information contained in the NEMS scenario descriptor (scedes) file.

the ICMM will actually solve. The user would then be able to visually view the properties associated with each sub-regional object for validation and verification purposes.

5.3 Preparing the Input Data

Python has an extensive set of libraries dealing with the reading and writing of data in various formats, data manipulation, and connection to various databases. Our recommendation is to use SQLite because it is already in use in the WEPS+ framework and the Open Database Connectivity (ODBC) drivers required by AIMMS. We will use Python to read in and prepare the model input data. The user-specified parameters will indicate the model regional aggregation level, and the data sets will be prepared accordingly.

5.4 Solving the Model

Once the data sets are prepared, Python will pass control to the AIMMS modeling software to generate the LP model and solve with a commercial solver, such as CPLEX (IBM ILOG CPLEX Optimizer). After solving, AIMMS will write data back to the SQLite database for interfacing with other models and reporting.

For each model run, the ICMM will solve each year of the forecast sequentially until all model years are solved. Once the ICMM has been solved for the entire forecast horizon, the results will be combined and reported back to the WEPS+ system. It should be noted here that this is not in conflict with each of the rest of the WEPS+ modules solving over the entire forecasting horizon at once instead of sequentially through time. Both approaches result in a forecast of the entire time horizon each time the models are called.

EIA's *International Energy Outlook* (IEO) is generally run in integrated mode (with all models on), but models can also be run individually, pointing to a common restart file.

The process of making integrated runs will not change with the new ICMM, and the ICMM will also provide the ability to run in standalone mode.

5.5 Aggregation and Reporting

When the ICMM cycle is finished, there will be a series of routines to collect the sub-regional output and to roll it up to WEPS+ regions (e.g., summing, averaging). A set of standard reporting tables will be produced from an output SQLite database at both the disaggregate level and at the WEPS+ 16 regions. This SQLite database would also be available to the user for ad hoc queries. This information could also be displayed in a GIS to allow the user to visually review the results. In addition, output coal model data will be aggregated (summing and/or averaging as is necessary) to the 16 IEO regions that are needed for the rest of the WEPS+ and placed into the binary restart file (including coal production, transportation, and prices).

6 Uncertainty and Limitations

There are many factors that introduce uncertainty in long-term forecasting models. Some of these may be analyzed by running alternative scenarios to gauge the sensitivity to a particular set of market conditions. Others can be mitigated over time with better data and continuous calibration.

Many exogenous assumptions need to be made in support of any model of the international coal market. These assumptions in many cases will drive the model solution results more than choices regarding regionality or the exact representation of production and transport. Examples of these assumptions are growth rates in countries like China and India, renewable technology characterizations, and regional climate change policies. Each of these uncertainties will bring to the surface the limitations of any given coal market forecast. The only way to address them is to run a number of scenarios across the spectrum of the key drivers.

Another key uncertainty is the availability and reliability of data that are used to feed the model. It is likely that the analyst will never have a complete picture of the global coal market and the behavior of all players involved. However, the data sets can be improved over time as new data become available.

Furthermore, there is some disagreement in the literature regarding the level of competitiveness in the global steam coal trade. While this trade has generally been considered to be competitive (Kolstad & Abbey, 1984), there have been indications in recent years of imperfect competition (market power) by certain players in the global coal market (Truby & Paulus, 2012). Specifically, the analysis in Truby et al. concluded that coal prices in the Asian market in 2006 were consistent with a competitive framework, but that prices and volumes in 2008 were not consistent with competitive version of their model. They infer from this that the level of competition in the Asian steam coal market may be decreasing.

Because of this, it may be difficult to have the model reproduce recent historical prices, and this therefore calls into question any “back-casting” exercise that may be performed to validate the model results. Back-casting can also be difficult due to the difference in future expectations in past years compared to what actually occurred (such as other fuel prices or economic growth). The design of the International Coal Market Module (ICMM) as a linear programming optimization model initially will leave open the possibility of experimenting with an MCP (mixed complementary programming) equilibrium modeling framework in the future.

The design of the ICMM described in this report will not only facilitate scenario analysis across the main market drivers, but it will allow greater depth in understanding regional impacts of such drivers. The design will also allow the analyst to have reasonable results in the absence of detailed regional data, with the possibility of adding modeling detail as more data become available. This could mean increased regional resolution or the separation of coal demands into a greater number of coal types. Finally, this design opens the door to investigating alternative modeling frameworks that address non-competitive behaviors.

7 Recommendations

The model design described in this document allows for a responsive staged-design approach. The suggested implementation process laid out here incorporates an evolutionary development schedule with delivery of products as early as possible, allowing EIA to provide feedback and suggest changes resulting in continuous improvements as the system is being built out. It is our recommendation that whomever is responsible for the design and development of the International Coal Market Module (ICMM) work in close coordination with EIA during each step of the process. With this in mind, the next steps would include:

- Discussions with EIA regarding the design
  - Software selection (SQLite, Python, selected Python libraries)
  - Optimization solver (Advanced Interactive Multidimensional Modeling System, AIMMS)
- Development of a process for capturing feedback/change request from EIA and incorporation in the design
- Development of schedule for rapid release of key prototype components
  - Front-end interface
    - Database design and implementation
    - User interface design and implementation
  - Main ICMM implementation
    - Data collection
    - Data storage and preparation routines
    - Optimization model
    - Back-end aggregation and reporting

The responsive design is implemented by swiftly producing a functional model to initiate a feedback loop with EIA and to identify changes that better suit one's needs. Figure 9 provides an overview of this process.

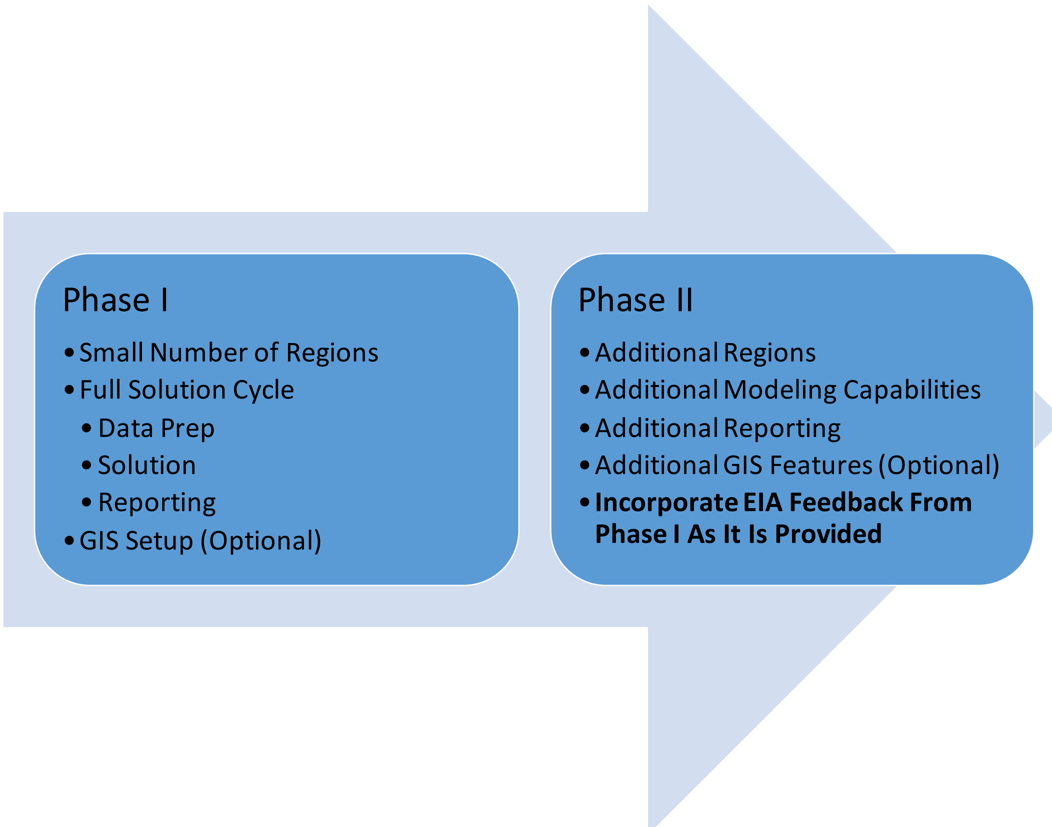

Figure 9: Overview of Responsive Design Process

7.1 Pre-development Software Design Review

Prior to the initial software prototyping phase, one should solicit feedback from EIA modelers regarding the proposed model design as part of a pre-phase initial design review process. This may involve suggestions regarding additional data sources, modifications to algorithms, addition of features, as well as solver selection criteria. A series of meetings that include detailed discussions to address specific design elements is our preferred path forward.

7.2 Phase I Prototype

The first phase of software development would involve building a stand-alone prototype in which all core design elements are constructed and tested. Specifically, the data structure of the ICMM will be created as well as a full solution cycle capability (data preparation -> linear programming (LP) matrix generation -> solver execution -> reporting). If EIA chooses to utilize a geographic information system (GIS) approach to manage spatial input data in the ICMM, then the GIS data system will be designed and implemented. Otherwise the database system will be created.

The prototype would focus on a subset of the 16 World Energy Projection System Plus (WEPS+) regions and would demonstrate the flexible design in creating appropriate sub-regions for the ICMM. An SQLite database would be constructed for the countries within that WEPS+ region that would be used by the Data Preparation Routine. The key elements performed by the routine will be developed, such as the ICMM regional definitions and associated data compilation (e.g., production capacity initialization, demand shares) and transportation infrastructure capacity. In addition, the framework for iterating from year-to-year and aggregating back to the WEPS+ region would be constructed.

In Phase I, a prototype for only one regional aggregation will be created. As these are being constructed and tested, experiments can be made with some of the choices of dimensionality.

The last step is to construct the reporting back end in which the ICMM region results are aggregated into the WEPS+ region results and formatted output results are constructed.

After the Phase I prototype has been completed, it will be tested with a set of alternative scenario assumptions to assess its robustness and response to a variety of market conditions.

At this point the prototype would be delivered and presented to EIA for review and comment. Work on Phase II will proceed while EIA digests the Phase I results and prepares a set of recommendations.

7.3 Phase II

Phase II of the ICMM development expands the Phase I prototype into the full model implementation. Here the feature set of the foundation built in Phase I will be expanded, as will the number of model regions in the solution set. These features could include additional reporting, automated verification of input data sets, and additional modeling capabilities (e.g., carbon taxes.) Recommendations from EIA, stemming from the Phase I prototype, should be incorporated at this time.

Additional GIS and user-input features relating to the visualization of the input data and the selection of certain model parameters may also be added during Phase II.

At this stage, a major portion of the effort is data gathering to complete the data set globally. In addition, a full benchmarking exercise will need to be performed on the EIA international energy balances for historical years.

Additional test scenarios will be added, at least some of which should include carbon emission taxes or penalty pricing. The purpose of these tests is only to verify the coal model's response, rather than to assess the full economic response that would occur when using the full WEPS+ framework.

As with Phase I, Phase II implementation will be delivered and presented to EIA for review and comment.

This phase could also include the development of an alternative mixed complementary programming approach for modeling non-competitive behaviors or that decision could be postponed until further experience with the full global model provides evidence of whether that approach would be desirable.

8 References

Aden, N., Fridley, D., & Zhang, N. (2009). *China's Coal: Demand, Constraints, and Externalities.* Lawrence Berkeley National Laboratory, LBNL-2334E.

Baruya, P. (2007). *Supply Costs for Internationally Traded Coal.* London, UK: IEA Clean Coal Centre.

Haftendon, C. (2012). *Economics of the Global Steam Coal Market - Modeling Trade, Competition and Climate Policies (Dissertation).* Berline: Technische Universitat Berlin.

Hellerworx, Inc. (2016). *DOE/EIA International Coal Supply Curve and Transportation Modeling Austrailia Pilot Study, Ocean Freight Forecast & Coal Data Survey Study Methodology and Documentation.*

Holz, F., Haftendorn, C., Mendelevitch, R., & von Hirschhausen, C. (2016). A Model of the International Steam Coal Market (COALMOD-World): Data Documentation 85. *DIW, https://www.diw.de/documents/publikationen/73/diw_01.c.546364.de/diw_datadoc_2016-085.pdf*.

IPCC. (2014). *Climate Change 2014: Synthesis Report.* Geneva: Intergovernmental Panel on Climate Change.

Kaptur, C. (2014, August). *An Optimal Model of Global Thermal Coal Trade.* Retrieved from Runge Pincock Minarco: http://www.rpmglobal.com/wp-content/uploads/2015/08/Issue124-Global-Thermal-Coal-Trade.pdf

Kolstad, C. D., & Abbey, D. S. (1984). The effect of market conduction on international steam coal trade. *European Economic Review, 24(1)*, 39-59.

Murphy, F., Pierru, A., & Smeers, Y. (2015). *Why Build Energy Models as MCPs? An Economic Perspective.* KAPSARC, KS-1513-SP07A.

OnLocation, Inc. (2014). *World Electricity Models: Final Report.* Prepared for EIA.

Paulus, M. (2012). *The Economics of International Coal Markets (Dissertation).* Energiewirtschaftliches Institut an der Universität zu Köln.

Truby, J., & Paulus, M. (2012). Market structure scenarios in international steam coal trade. *The Energy Journal 33 (3)*, 91-123.

U.S. Energy Information Administration. (2016). *International Energy Outlook 2016.* U.S. Energy Information Administration.

US Energy Information Administration. (2014). *Coal Market Module of the National Energy Modeling System: Model Documentation.* EIA.

US Energy Information Administration. (2016, February 21). Retrieved from EIA Open Data API: https://www.eia.gov/opendata/qb.php?category=724255&sdid=COAL.IMPORT_QTY.TOT-CO-BO_MA.A

Appendix: Details of Net Present Value Approach

The Net Present Value (NPV) of a stream of values *R(t)* is defined as the sum of the present values of the individual cash flows:

$$NPV_{total}\left[R(t), d\right] = \sum_{t} R(t) * (1+d)^{-t}$$

where *d* is the discount rate. This is what we refer to as the **NPV total** and would be used for all *costs and prices* in the ICMM.

We define the **NPV average** of a stream of values to be the NPV of those values divided by the NPV of time:

$$NPV_{avg}\left[R(t), d\right] = \frac{\sum_{t} R(t) * (1+d)^{-t}}{\sum_{t} 1 * (1+d)^{-t}}$$

The NPV average would be used for any *quantity-related* coefficients in the ICMM.

The main idea of using these two different definitions of NPV calculations is to enable the model to represent the total cost/revenue of an activity in planning periods that are comprised of multiple forecasting years. For a given decision vector *X(t)* and cost vector *C(t)*, assume that the total objective function value of operating this activity in all years contained in a planning period *p* would be:

$$OBJ = \sum_{t \in p} C(t) * X(t) * (1+d)^{-t}$$

However, a planning period containing multiple forecast years is represented as a combined single "lump" and so we are not choosing each *X(t)* in the model, only a combined decision vector *X(p)* representing all of the years contained in the planning period *p*. We therefore assume that *X(p)* represents the average activity over all years in a given planning period. The objective function then becomes:

$$OBJ = X(p) * \sum_{t \in p} C(t) * (1+d)^{-t}$$

Assume further that *X(t)* is subject to a constraint in the model:

$$X(t) \geq D(t), \qquad \forall \; t \in p$$

where $D(t)$ is a vector of scalar right-hand side values. Then we want $X(p)$ to be subject to an average constraint:

$$X(p) \geq \frac{\sum_{t \in p} D(t) * (1+d)^{-t}}{\sum_{t \in p} (1+d)^{-t}}$$

The right-hand side coefficient is now a discount rate-weighted average in which near-term quantities have a heavier weighting than quantities in later years.

Example of an NPV average calculation:

Assume that a planning period represents three model years. Let $R(t)$ = (1.0, 1.5, 2.0) for t = (0, 1, 2) and d = 0.1. Then the NPV average of $R(t)$ is calculated as the following:

$$NPV_{total}[R(t), d] = \frac{R(0)}{(1+0.1)^0} + \frac{R(1)}{(1+0.1)^1} + \frac{R(2)}{(1+0.1)^2}$$

$$= \frac{1}{1.1^0} + \frac{1.5}{1.1^1} + \frac{2.0}{1.1^2} = 4.017$$

$$NPV_{time}[1, d] = \frac{1}{(1+0.1)^0} + \frac{1}{(1+0.1)^1} + \frac{1}{(1+0.1)^2} = 2.736$$

$$NPV_{avg}[R(t), d] = \frac{NPV_{total}[R(t), d]}{NPV_{time}[1, d]} = \frac{4.017}{2.736} = 1.468$$