Concepts, Data Sources, and Techniques
Handbook of Energy Modeling Methods

World Energy Projection System (WEPS): Hydrocarbon Supply Model

Independent Statistics & Analysis
www.eia.gov

U.S. Department of Energy
Washington, DC 20585
Last Updated 2020

This report was prepared by the U.S. Energy Information Administration (EIA), the statistical and analytical agency within the U.S. Department of Energy. By law, EIA's data, analyses, and forecasts are independent of approval by any other officer or employee of the United States Government. The views in this report therefore should not be construed as representing those of the U.S. Department of Energy or other federal agencies.

Contents

Table of Tables

Table of Figures

1. Introduction

The WEPS Global Hydrocarbon Supply Model (GHySMo) projects the global supplies and prices of liquid and gas hydrocarbon fuels. It comprises four models:

- The Upstream Model simulates the activity of the discovery and production of primary hydrocarbon resources (several types of crude oils, natural gas, and hydrocarbon gas liquids) at various locations.
- The Conversion Model projects the behavior of global refining centers in transforming various types of crude oils into refined petroleum products.
- The Logistics Model simulates the transport of both primary hydrocarbon commodities (for example, crude oil) and refined petroleum products (for example, gasoline) between the locations represented in the Upstream and Conversion Models. It also simulate the transport of refined products to end-use demand locations (currently the 16 WEPS regions).
- The Integration Model controls the iteration of the other models (or reduced forms of them), performs pre- and post-iteration steps, checks for convergence, and reports projections to the WEPS.

Although GHySMo was designed to be flexible in the exact fuels that it can represent, GHySMo models the following commodities in all of its models:

- Three grades of crude oil (light, medium, and heavy)
- Condensate
- Natural gas
- Mixed hydrocarbon gas liquids (HGLs)
- Motor gasoline
- Conventional blendstock for oxygenate blending (CBOB)
- Diesel fuel
- Heating oil
- Mixed kerosene/jet fuel
- Residual fuel oil
- Sequestered petroleum products
- Other petroleum products
- Biofuels

2. Integration Model

2.1 Introduction

The GHySMo Integration Model controls the running of the GHySMo Logistics and Conversion Models. It transfers data among the GHySMo models and between GHySMo and WEPS, and it determines whether or not GHySMo's linear program solutions have converged. The GHySMo Integration Model iterates among the Logistics Model, the Conversion Model, a reduced-form version of WEPS demand models, and a reduced-form version of the Upstream Model until the resulting values at the end of the cycle are close enough to the values at the beginning of the cycle to meet convergence criteria, or until a maximum number of iterations is reached. Before it can be used as part of a WEPS run, GHySMo must first be run in a stand-alone mode. The stand-alone run initializes the GHySMo models so that they are able to efficiently solve their respective linear programs when used in the context of the WEPS.

2.2 Pre-iteration-cycle calculations

Whether it is being run in stand-alone mode or as part of a WEPS run, the GHySMo Integration Model must first run a set of pre-processing steps before it enters its iterative loop. When run in stand-alone mode, the model reads in initial data from the Upstream and Conversion Models. From the Upstream Model, it reads the base supply quantities and prices for the commodities represented. From the Conversion Model, it reads in the base crude oil demand quantities and prices by crude oil type and the base supply quantities and prices for refined petroleum products.

Reaggregating product quantities demanded

The Integration Model receives, from the WEPS demand models, the projected product demands for all commodities for which GHySMo will be projecting prices and quantities supplied. It aggregates the demanded quantities for each commodity in each WEPS region. Because GHySMo does not operate on exactly the same petroleum product categories represented in the WEPS, some WEPS products must be separated into one or more GHySMo commodities, and other WEPS products are combined into a single GHySMo commodity. For example, WEPS only represents distillate fuels as one category, but GHySMo separately represents diesel fuel and heating oil, so the WEPS distillate commodity stream must be split into separate diesel and heating oil streams. On the other hand, WEPS considers kerosene and jet fuel to be distinct commodities, whereas GHySMo treats them as a single commodity.

STEO benchmarking

The Integration Model performs ratio adjustments to ensure consistency between GHySMo's initial prices and quantities and the U.S. Energy Information Administration's (EIA) short-term projections published in the *Short-Term Energy Outlook* (STEO). For the first one to two years of the projection period, the model adjusts projections of crude oil, lease condensate, and natural gas plant liquids production to agree with STEO production forecasts. It also adjusts projections of crude oil prices to agree with STEO forecasts of the Brent crude oil price.

To avoid an abrupt change in the projections for the year following the final STEO year, EIA analysts specify an adjustment phase-out period. During the phase-out period, the projections are a weighted average of the adjusted and unadjusted GHySMo projections; the weight assigned to the unadjusted projection gradually increases to reach unity at the end of the period.

Accounting for contract liquefied natural gas

The final step in the pre-processing is to account for fixed-destination liquefied natural gas (LNG) contract sales. For any LNG contract sales for which EIA knows the source and destination, the Integration Model subtracts the contracted LNG flow from the source location's supply and the destination location's demand. The Integration Model also keeps track of any natural gas used for liquefaction or regasification for these fixed-destination LNG flows.

2.3 Iteration cycle

The main function of the GHySMo Integration Model is to cycle between the various other GHySMo models (or reduced forms of them) until the prices and quantities of the commodities reach an equilibrium. This function is similar to the function the main WEPS Convergence Model performs within the WEPS system. Normally, the Integration Model begins its iteration cycle with the Logistics Model. However, in stand-alone mode, it runs the Conversion Model first. The Integration Model initializes the Logistics Model with the base prices, base quantities, and elasticities for both supply and demand at all locations represented.

After the Logistics Model has executed, the Integration Model compares the prices returned by the Logistics Model to the prices input to the Logistics Model to determine if they are within specified tolerances for convergence. After checking price convergence, the Integration Model runs the Conversion Model, using the Logistics Model's results to update the projected natural gas prices at the natural gas processing locations represented.

Reduced-Form Demand Model

After the Conversion Model has executed, the Integration Model then executes the Reduced-Form Demand Model. This model uses simple isoelastic demand curves (based on the elasticities estimated by the Logistics Model) to estimate the response of the WEPS demand models to the new prices generated by the Logistics Model. Because GHySMo iterates between its various models before returning projections to WEPS, this simplified version of the WEPS demand models allows GHySMo some ability to anticipate changes in the WEPS projected demands during the GHySMo iteration process.

Reduced-Form Upstream Model

Because the GHySMo Upstream Model cannot be called directly from the GHySMo Integration Model, the Integration Model executes the Reduced-Form Upstream Model after the Reduced-Form Demand Model. Similar to the Reduced-Form Demand Model, the Reduced-Form Upstream Model uses isoelastic supply curves (based on the elasticities from the Logistics Modell) to estimate the response of the Upstream Model to the new prices generated by the Logistics Model. After executing the Reduced-Form Upstream Model, if the prices had already converged after the execution of the Logistics Model, the Integration Model compares the quantities input to the Logistics Model against those output from the other models to evaluate convergence of the quantities.

Checking for convergence

The GHySMo Integration Model can use different, although similar, algorithms to check for convergence of prices and quantities. We describe the default algorithm here. All the algorithms compare the final post-iteration set of supply and demand price and quantity projections, by location, commodity, and year, with the corresponding set of projections initially input to the Logistics Model, to determine whether or not the average absolute percentage differences are within the convergence tolerance limits.

When checking convergence of quantities, the Integration Model calculates, for each commodity, the mean absolute percentage difference between the final and initial values across all locations and years. However, when checking convergence of prices, for each commodity, the Integration Model calculates the *quantity-weighted* average absolute percentage difference between the final and initial values across all locations and years. The absolute percentage differences are then averaged over a specified number of iterations before testing for convergence. If the average for a commodity is at most the specified tolerance, then the corresponding price or quantity is considered to have converged. If either prices or quantities have not converged for all commodities, the iteration loop begins again, and the resulting final values are used as the new inputs to the Logistics Model.

2.4 Post-iteration-cycle calculations

After the prices and quantities have converged or the Integration Model has executed the specified maximum number of iterations without convergence, the Integration Model restores the volumes of fixed-destination LNG contracts to the appropriate supply and demand locations. The Integration Model then performs the STEO benchmarking again to ensure that the final GHySMo price and quantity projections agree with the STEO forecasts for the STEO years.

Projecting sectoral prices

The Integration Model then applies predetermined sectoral markup factors to the GHySMo wholesale prices to project sectoral prices for the industrial, commercial, transportation, and residential sectors. Prices are first marked up to industrial sector prices, which may have a cap as to the maximum amount prices can change between years. After this adjustment, prices are marked up from the industrial price to the other sectoral prices, by either adding or multiplying by predetermined adjustment factors. To avoid abrupt price changes, prices may be smoothed across years, depending on options set by EIA analysts.

Projecting refinery and natural gas plant fuel use

The Integration Model also computes the refinery fuel use and refinery processing gain. It projects the use of natural gas, still gas (counted as *other petroleum*), catalyst coke (counted as *petroleum coke*), electricity, and steam (heat). Although the Conversion Model provides these projections, the Integration Model aggregates them from the Conversion region level to the level of the WEPS regions. To project refinery gain, the Integration Model subtracts the volumes of the crude oils demanded (including lease condensate) from the volumes of the products supplied by the refinery, including still gas and catalyst coke, but excluding biofuels.

The Integration Model then projects the natural gas lease and plant fuel use, the natural gas used in LNG transport, and the natural gas used in pipeline transport. It assumes that natural gas lease and plant fuel use is a constant fraction of dry gas production, and it reports the projected values to the WEPS Industrial Demand Model. The Logistics Model accounts for natural gas used in LNG and in pipeline transport, except for contracted flows that the Integration Model removes from the linear program representation before beginning the iteration cycle. The Integration Model adds any natural gas used in fixed-destination LNG contract processing to the value reported by the Logistics Model before passing the projections to the Industrial Demand Model. The Integration Model passes the projected quantities of natural gas used in pipeline transport to the Transportation Demand Model of WEPS. Finally, the Integration Model reports the projected quantities of natural gas and crude oil produced by the Upstream Model.

3. Upstream Model

3.1 Introduction

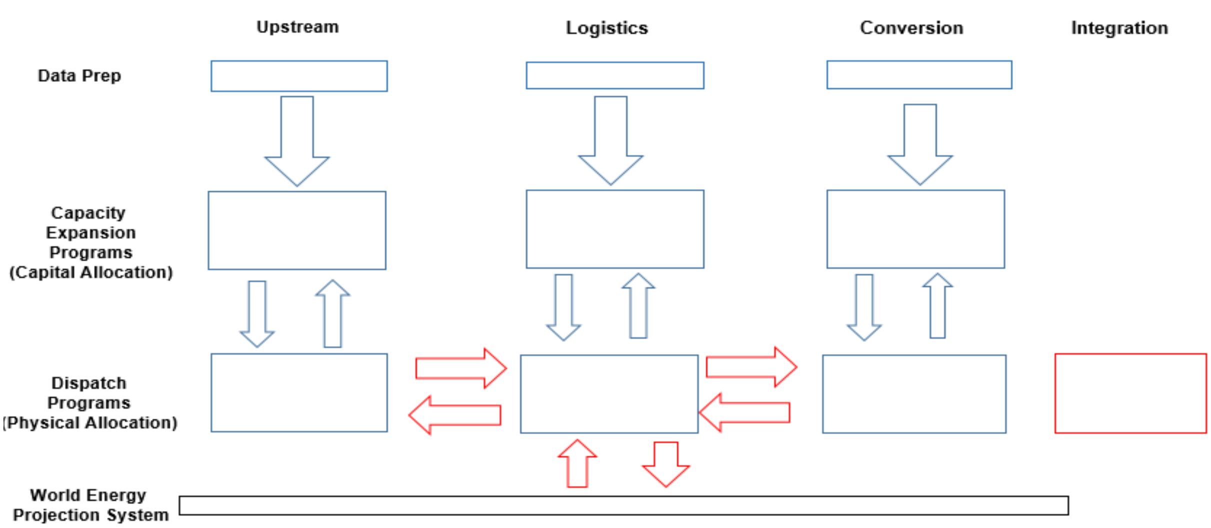
The GHySMo Upstream Model represents the upstream exploration, drilling, and production sectors of the international hydrocarbon economy. It simulates the development of hydrocarbon resources and production operations over lifetimes of oil and natural gas fields and wells, assuming that field- or well-level production will decline over many years until closure. The model projects production for each resource and then adds the projections for regional groups defined by EIA analysts. It assumes that the most economical resources will be developed first and schedules drilling rigs competitively for oil and natural gas projects. Oil and natural gas composition and annual production profiles are specific to national or regional-level resource and development assumptions.

The Upstream Model draws on many international datasets to estimate annual capital and operating expenses needed for hydrocarbon production, available numbers of drilling rigs, well productivity, oil and natural gas composition of wells, and applicable taxes and royalties. In the upstream portion of the oil and natural gas industry, the expenses, production rates, available drill rigs, and taxes fluctuate significantly with commodity prices. Within each region, for each projection year, new resource parcels are projected to be developed for production based on rankings of net present values per drill rig year. The model generates annual regional projections of oil and natural gas production based on the expected commodity prices. The production rates are passed to other GHySMo models (such as Logistics and Conversion). The Upstream Model results rely on input data from many sources, including the other GHySMo models.

The Upstream Model was designed to be highly flexible in representing known (or speculative) oil and natural gas resource types and simulating economic competition between global resources. The Upstream Model can function with minimally necessary information or with extensive historical data. It is based on a thorough technical understanding of crude oil and natural gas resources and how their development responds to prices. In coordination with Integrated GHySMo and WEPS, it projects supply under all global demand scenarios and commodity price combinations.

3.2 Relationship to Integrated GHySMo and WEPS

Figure 1. Components of GHySMo



Information flows into and out of the Upstream Model as follows:

- The Logistics Model and Integration Model provide the Upstream Model with the following:
    - Annual unit prices for natural gas fractions (dollars per thousand cubic feet) and crude oil fractions (dollars per barrel) in each supply region that reflect price *after* logistics costs have been subtracted out
    - Annual Brent crude oil prices (used as a basis for computing costs and production rates)
- The Upstream Model provides the following to the Logistics Model and Integration Model, by supply region
    - Annual production of crude oil and natural gas products
    - Annual production as a result of newly developed resources
    - Annual production by designated gas or oil development region

3.3 Modeling methods

The Upstream Model can primarily be described as an international reserve and resource calculator, divided into four submodels processing information sequentially. The submodels process resource estimate information from distinctly different types of geologic and economically productive environments in four different resource information categories.

The four submodel calculators of the Upstream Model are:

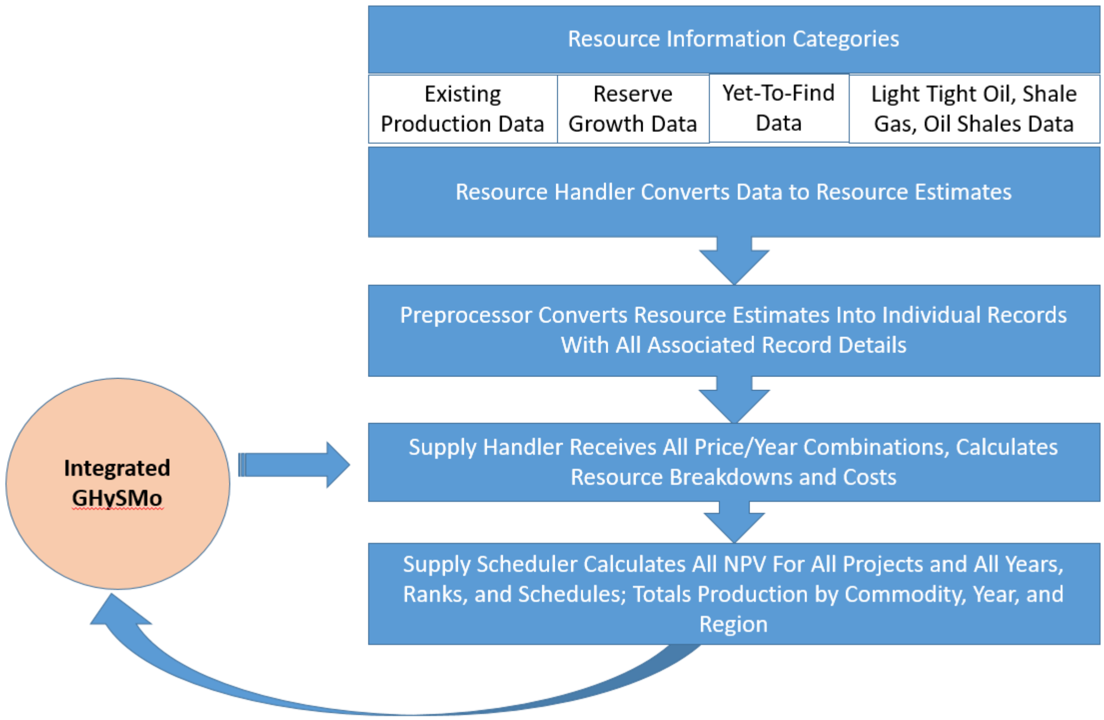
- Resource Handler (RH)
- Preprocessor (PP)
- Supply Handler (SH)
- Supply Scheduler (SS)

The four resource information categories accounted for within each submodel are:

- EP—Existing crude oil and natural gas production, analogous to reserves
- RG—Reserve growth of existing crude oil and natural gas fields, analogous to resources
- YF—Yet-to-find crude oil and natural gas resources, resources expected to be found based on past discovery patterns and resource volumes
- UC—Resources not accounted for in the previous three categories, such as light, tight oil and shale gas, tar sands, methane hydrates, or oil shales such as those found in the Rocky Mountain basins of the United States (sometimes referred to as unconventional resources)

The four calculator submodels process stochastic resource calculations, consolidate resource estimates into common formats, collect Integrated GHySMo generated price information, combine prices and costs at various levels, and process the scheduling of resource development across all four resource information categories for the given price path. Figure 2 is a flow chart of this process.

Figure 2. Flowchart of internal Upstream Modelupstream operation



3.4 Model and submodel descriptions

This section of the documentation describes the main working procedures, processes, and calculations of the four main working submodels of the Upstream Model.

Resource handler submodel

The Resource Handler (RH) submodel accepts raw probabilistic resource assessment inputs and converts them into structured and predetermined deterministic resource estimates. The calculations are designed to accept raw input data from four distinct resource information categories and convert them into similar forms that are then used as input data for downstream submodels. Each of these four resource information categories have specific types of calculations necessary to create a common resource estimate output. The individual, smallest level of resource record in the RH is a resource block. A resource block contains basic resource size and related information and is the smallest common unit of information within the RH. The basic equations used are described below within each resource information category for creating these resource blocks.

Resource Handler description of equation groups

EP and RG integrated calculations

EP data most closely resemble a reserve estimate of crude oil and natural gas resources. Reserve estimates are highly certain and are well defined by industry. RG information is analogous to resource estimates within existing fields of volumes that have been discovered within existing oil and gas accumulations but are not yet producing. They are generally less costly to develop; they require no finding or exploration costs, the properties of the rock and fluids are known, and they can take advantage of existing infrastructure. EP and RG calculations are linked and made in sequence because EP outputs are part of the input data needed by the RG calculation.

The primary EP calculation begins from a known time and producing rate of crude oil or natural gas, and it extrapolates an additional volume of crude oil and natural gas out into the future with a given rate of decline. This value is analogous to a current future reserves number. Cumulative production to date and this reserve number estimated in the future are the current known recoverable volume from a given resource block. The calculated EP reserve estimate is the record-level EP resource number.

The EP resource number estimate is then dedcted from the original in-place estimate of that resource, leaving a total remainder available for development in the future. This total remainder is the absolute maximum amount of reserve growth possible, and as such it is not expected to be the amount extracted in the future. To determine the amount extracted in the future, the U.S. Geological Survey (USGS) reserve growth function for world reserve growth[1] is used to scale back the absolute maximum number. This scaled back answer is the RG resource number estimate.

It should be noted that the EP-RG process is stochastic in nature, and the description above has been simplified to the procedure itself. The uncertainty built into the EP estimate is the decline rate, which is described by a triangular distribution. The uncertainty built into the RG calculation is related to the fraction of the total original in-place resource attributed to cumulative production and expected reserves assumed to have been removed by the EP calculation. The higher the fraction of original in-place resource that has been produced, the more the distribution provided by the USGS is truncated when the RG fraction remaining calculation is made, and the less overall reserve growth is available for future development.

YF calculations

Yet-to-find (YF) volumes, often referred to as undiscovered, are those volumes presumed to exist based on the discovery process modeling of undiscovered resources,[2] in this case the detailed work released by the USGS in the 2012 update of the 2000 World Assessment.[3]

USGS published work and procedures were used to subdivide assessment distribution results into field size classes and their corresponding average size and quantity estimates. Ancillary data included in the USGS results were matched to the corresponding field size and number estimates.

As with the EP and RG calculations, YF calculations also function for given input probabilities. USGS resource estimates of undiscovered resources are themselves probability density functions, from which

[1] U.S. Geological Survey Assessment of Reserve Growth Outside of the United States, 2015: https://pubs.usgs.gov/sir/2015/5091/sir20155091.pdf

[2] U.S. Geological Survey Assessment Concepts for Conventional Petroleum Accumulations (2005): https://pubs.usgs.gov/dds/dds-069/dds-069-d/REPORTS/69_D_CH_19.pdf

[3] Supporting Data for the U.S. Geological Survey 2012 World Assessment of Undiscovered Oil and Gas Resources (2013): https://pubs.er.usgs.gov/publication/ds69FF

any given probability of resource size can be calculated. When combined with the ancillary information, the YF resource estimates ultimately arrive in the same format as the EP and RG outputs.

UC calculations

Global Unconventional (UC) resource estimates included within the Upstream Model are derived primarily from the EIA World Shale Resource Assessments study.[4] This information provided resource estimates of recoverable tight oil and shale gas in select countries and basins around the world. However, the experience in developing these resources at scale and in terms of the distribution of outcomes and costs exists primarily in the United States. To create costs and distributions of results for these international plays, production analogs were identified within the United States, and this information was applied with respect to development cost and the distribution of results per unit area internationally, based on the EIA World Shale Resource Assessments study.

The results of the RH calculations are that all raw data resource estimates and related ancillary data for a given probability are converted into complete resource blocks, in the same way that the outputs from the other resource information categories are converted, namely that each one represents a well-defined group of crude oil or natural gas resources available in the future.

Preprocessor submodel

The Preprocessor (PP) is responsible for converting RH derived resource estimates into complete detail records. The RH outputs these resource estimates as a resource block, which is the smallest level within the Upstream Model system. The PP converts these resource blocks into complete records called cost blocks. These cost blocks are a combination of RH-generated resource estimates and detailed additional information such as related costs, resource scalars, and modeled coefficients of other processes and data, all related to the particular characteristics of a single cost block. External data sources, additional analysis, and outputs from other calculations are all used in the construction of cost blocks. Cost blocks are the basic record used by the rest of the Upstream Model systems.

Preprocessor—description of equation groups

The PP uses the RH outputs and additional information (detailed in the Model Input Data and Sources section) to generate the following corresponding detail for each cost block:

- Gas-to-oil (GOR) ratio
- Natural gas liquids (NGL) factor
- API gravity
- Sulfur content
- Oil supply designation based on API gravity and sulfur content
- Average depth of formation
- Water depth
- Natural gas composition
- Scalars to describe resource size probabilities

[4] EIA World Shale Resource Assessments (2014): https://www.eia.gov/analysis/studies/worldshalegas/

Using the basic initial detail of a cost block, the PP uses the particulars within each cost block in combination with additional sources of information to calculate the following for each cost block:

- The rig numbers and types required for development of a given resource are based on the resource size, drilling, and water depth
- The composition of the production profile parameters for both crude oil and natural gas that is based on resource size and coefficients for rig efficiency and type
- First years available for production scheduling varies by original resource information category
- Operating expenses including national level taxes using a study from an external data source, resource size, and onshore or offshore characteristics
- Wildcat wells and timing of new discoveries based on search coefficients, undiscovered size, and quantities of resource

Supply Handler submodel

The Supply Handler (SH) accepts all cost block records from the PP, and it begins the interaction with Integrated GHySMo. The primary purpose of the SH is to begin the process of creating all records split by commodity and time to match to time variable cost parameters for all given commodities by price and year. The SH begins by inputting the full price path from Integrated GHySMo for all commodities for all years.

Supply Handler—description of equation groups

The SH creates a set of extended records generated from cost blocks that cover production starting in any year for different cost points, computing time variable operating expenses and taxes, production volumes, and revenue, all through a given amount of time. User designated parameters determine the minimum size of a cost block to be handled individually or blended together to make a continuous amount of resources from multiple cost blocks. As opposed to equations, the primary purpose of the SH is splitting, filtering, grouping, and aggregating existing data into a form amenable to economic ranking and scheduling.

Supply Scheduler submodel

The Supply Scheduler (SS) submodel is responsible for calculating the economic ranking, scheduling, and output of all resource development.

Supply Scheduler—description of equation groups

The primary equations used within the SS are focused on calculating the net present-value of all projects. After the economic ranking, projects are deployed with the condition that rigs are available to develop the project. These results are filtered and totaled for each year to generate all commodity outputs by time and volume. These volume outputs correspond with the original time, commodity, and price of the original input price path.

3.5. Model input data and sources

Resource Handler

Enverus DrillingInfo

- Well-level data for all U.S. light, tight oil, and shale gas plays

U.S. Energy Information Administration

- Grid Based Assessment and Cost Calculator (GBACC) to generate resource supply curves from U.S. well-level production data
- 2014 World Shale Resource Assessment study
- Analysis of geologic similarities between formations contained within the 2014 World Shale Resource Assessment and U.S. analog formations
- International oil field decline rate study of IHS EDIN oilfield data by field size, age, and onshore or offshore location
- EIA analysis and estimates of rig-type requirements, drilling efficiency, and costs for development of light/tight oil and shale gas internationally

U.S. Geological Survey

- Reserve growth studies, methodology, and international crude oil and natural gas functions (previously referenced)
- Supporting data for the U.S. Geological Survey 2012 World Assessment of Undiscovered Oil and Gas Resources (previously referenced)
- Field size class designations for crude oil and natural gas

RyStad UCube

- International capital cost data for oil and natural gas resources by size
- Analysis and parameterization of capital costs by USGS field size classes

IHS Markit—EDIN Database

- Aggregated international resource volumes and details

Preprocessor

U.S. Energy Information Administration

- EIA National Energy Modeling System (NEMS) Oil and Gas Supply Model discovery process procedures and coefficients, based on USGS discovery process modeling principles and EIA Gulf of Mexico analogs
- Crude oil and natural gas production ramp years and profile from the Oil and Gas Supply Model of NEMS

Rystad UCube

- Tax, royalty, fixed, and variable operating cost data
- Analysis and parameterization of tax, royalty, fixed, and variable operating cost data

Baker Hughes Rig Data

- Historical international rig counts, by region and rig type
- EIA analysis of regional historical rig type and count in relation to crude oil and natural gas prices, and parameterized adjustments to regional rig numbers/productivities based on natural gas prices, allowing additional rigs for scenarios of higher natural gas prices

Supply Handler

Oxford International Discount Rates

- Country-level capital risk expectations used for net present-value discounting calculations

Bloomberg Historical Crude Oil Benchmark Pricing

- Data on historical differences between Brent crude oil prices and other common crude oil benchmarks

3.5.4. Supply Scheduler

U.S. Energy Information Administration

- Most recent *Annual Energy Outlook* results for U.S. crude oil and natural gas production for Reference Case and Low and High Oil Price Cases

4. Conversion Model

4.1 Introduction

The GHySMo Conversion Model (Conversion) models the downstream refining sector of the international hydrocarbon economy by simulating (as a linear program, LP) the operation of all international petroleum refineries. These international refineries are aggregated into three representative world Conversion regions that center on the following locations:

- Europe
- Asia
- U.S. Gulf Coast

The refining sector's function is to convert crude oil into consumer petroleum products, such as gasoline, diesel fuel, jet fuel, propane, heating oil, and residual oil. The LP simulates the selection and conversion of various crude oil types (based on API gravity: very light (condensate), light, medium, and heavy) into these petroleum products, assuming that the refiner's goal is to maximize profits within each Conversion region for each projection year, subject to a number of constraints. The model generates projections of the following key information related to liquid fuels production:

- Energy prices for petroleum products and biofuels (at the refinery)
- Volumes of crude oil needed to produce petroleum products
- Volumes of fuel consumed (natural gas, refinery gas, catalyst coke), byproducts (sulfur), and utility demands (electricity and steam)

This information is passed to other GHySMo models for further processing. For example, projected volumes of crude oil needed to produce petroleum products in each Conversion region are passed to the Logistics Model, which simulates the interregional movement of crude oil from upstream production regions to Conversion regions to meet these projected crude oil demands. For each Conversion region and projection year, the Logistics Model provides the following information to the Conversion Model:

- New energy demands and corresponding expected prices by product type
- Crude oil prices and corresponding expected demands by crude oil type
- Utility prices (electricity and steam)
- Biofuels demands (combined ethanol and biodiesel)

This cycle continues until the Logistics Model determines that all models in GHySMo agree on the same optimal solution: a set of values that define

- Crude oil production and consumption volumes and price

- Corresponding production of petroleum liquids from the crude oil to meet product demands (as defined by the WEPS demand models)
- Resulting liquid fuels product prices based on regional production and transport costs

4.2 LP objective function

The Conversion Model maximizes an objective function that represents the sum of all components of refinery profit (revenue minus cost) for each Conversion region and projection year. Refinery profit comprises the following revenue and cost components:

- Revenue from product sales
- Crude oil and other feedstock prices (costs)
- Fixed costs incurred for maintaining and operating existing refinery capacity
- Costs associated with building new refinery capacity
- Utility costs associated with operating the refinery (electricity and steam)
- Miscellaneous operating costs specific to refinery operating modes (for example, cost for extra Fluid Catalytic Cracking (FCC) unit catalyst needed for an operating mode that maximizes gasoline production)

Transportation costs associated with transporting crude oil to each Conversion region are not explicitly modeled by the Conversion Model. These costs are included in the regional crude oil prices that the Logistics Model provides to the Conversion Model.

4.3 LP decision variables

For each Conversion region and projection year, the Conversion Model solves for the following decision variables that contribute to maximizing profit. More details on type, content, and descriptions are provided in the Refinery representation section later in this report.

- Selection of operating levels for each operating mode associated with each refinery type
- Selection of crude oil blend options during operation of each refinery type
- Amount of crude oil purchased, by type, during operation of each refinery type
- Amount of products produced, by type, during operation of each refinery type
- Amount of intermediate stream blended into product
- Amount of energy (electricity, natural gas, steam) required during operation of each refinery type
- Amount to increase or reduce refinery unit capacity (atmospheric distillation units (ADU) and downstream upgrading units

4.4 Key LP constraints

The profit maximization is subject to four key constraint categories. More details on type, content, and descriptions are provided in the Refinery representation section.

Demand satisfaction

For each Conversion region, projection year, and product type

- The amount of product produced (from either intermediate streams or ADU yields) plus distress imports must equal total demand for that product (represented by a demand curve) plus distress exports.

- Distress imports and exports are activated in the LP only when the product produced cannot meet the totals demand for that product, indicating a shortfall or excess of products produced (respectively).

Crude oil supply

For each Conversion region, projection year, and crude oil type

- Total refinery consumption of a crude oil type must equal total supply of that crude oil type.
- Crude oil supply is represented in the LP by an isoelastic supply curve, constructed from
- The crude oil price provided by the Logistics Model
- The quantity demanded in the previous GHySMo cycle
- A supply elasticity (analyst judgement)

Utility balance

For each Conversion region, projection year, and utility type (electricity, steam)

- The amount of utility purchased must equal the sum (over refinery type and operating mode) of the utility consumed for each operating level plus the sum (over downstream unit) of the utility consumed to convert each intermediate stream into petroleum product

Capacity balance and expansion

For each Conversion region, projection year, and refinery type:

- Operation level cannot exceed the sum of existing and new capacity less retired capacity.
- Simple refineries (see Refinery Representation section): The model assumes no unplanned builds.

For each Conversion region, projection year, and downstream upgrading unit type:

- Operation level cannot exceed the sum of existing and new capacity (if allowed) less retired capacity.
- Secondary upgrade units (see Refinery representation section): The model assumes no unplanned builds.

4.5 Refinery representation

In addition to some background on how the LP was formulated in the Conversion Model, this section provides more details and context about the various refinery terms and classifications presented in the sections above:

- Refinery types
- Primary and secondary downstream upgrading units
- Operating modes
- Crude oil types and crude oil blends
- Petroleum products
- Intermediate streams

Three refinery types are represented in the Conversion Model:

- Cracking (non-simple, with optional coking)
- Hydrocracking (non-simple, no coking)

- Topping/hydroskimming (simple)

However, unlike many refinery models, the refinery representation in the Conversion Model does not explicitly model each downstream process unit that exists for these three refinery types. Instead, it uses intermediate stream and product yield data associated with a set of operating modes to represent the conversion of crude oil to petroleum products. These operating modes are defined using a detailed engineering model (GRTMPS, Generalized Refining Transportation Marketing Planning System, by Haverly Systems, Inc.). This approach converts the complex operations of each refinery type into a lean representation in the LP model. However, as an enhancement, downstream upgrading units were added to capture the extra costs associated with producing certain products from intermediate streams. The three primary downstream upgrading units associated only with the operation of the two non-simple refinery types are cokers, crackers, and hydrocrackers. These key refinery processing units break down the heavier gas oil streams (long carbon chains) into naphtha and distillate intermediate streams and typically produce petroleum coke. In addition, the following three secondary downstream units were added to all refinery types to capture corresponding cost and capacity build decisions in the LP: reformers, naphtha hydrotreaters, and distillate hydrotreaters. These units are key to achieving the quality specifications required for some products (such as octane levels in gasoline or cetane levels in diesel).

As identified above, GRTMPS was used to develop the operating modes and their corresponding yield and utility-consumption parameters associated with selected crude oil blends for each of the three refinery types. These data are provided as input to the LP. The main advantages to developing a slate of refinery operating modes from a detailed, complex, and robust refinery model (GRTMPS) are that it

- Promotes confidence in its simplified representation
- Enforces a material balance (in other words, ensures petroleum products are created from the necessary refinery inputs)
- Maintains control of distillation unit cuts into intermediate streams and products (in other words, liquefied petroleum gas [LPG] cannot become diesel fuel)
- Provides a direct accounting of refinery fuel use and utility requirements that are then reported to WEPS

The yield data for each operating mode are grouped into two categories: data on products and data on intermediate streams (for example, refinery streams that will become finished products). By representing intermediate streams, the Conversion Model accounts for the fact that some refinery liquid streams, such as middle distillates and naphtha, may be further processed (at a cost) into higher value products, such as diesel or CBOB (conventional blendstock for oxygenate blending) to produce finished gasoline for final sale, respectively. The same intermediate stream could also undergo more limited (less costly) processing into lower value products, such as residual fuel oil (resid), other petroleum (aviation gasoline, naphtha and other oils for feedstocks, miscellaneous products), or sequestered petroleum (asphalt and road oil, lubricants, waxes). (For general information on middle distillates, residual fuel oil, naphtha, and other intermediate streams produced at a refinery, refer to EIA's Energy Explained: Refining crude oil.) Table 1 presents the products and intermediate stream-to-products conversions represented in the LP model.

Table 1. Operating mode yield options—directly to products or by intermediate streams

	Secondary intermediate stream	Products*								
	Kerosene	LPG	Gasoline	CBOB	Diesel	Heating oil	KeroJet	Resid	Petroleum _Other	Petroleum _SEQ
Direct to product		x				X	x	x	x	x
Intermediate streams to product										
Naphtha			x	x					x	
Distillate	x				x	X				
Residuum								x	x	x
Kerosene							x			

* Product definitions are in the Abbreviations and Acronyms section.

4.6 Crude oil blends

Crude oil blends selected by each refinery can consist of a single crude oil type, or any combination of the four crude oil types represented in the model: light, medium, heavy, and condensate. (For more on crude oil types, see Energy Explained.) For many operating modes, more than one crude oil blend is available to produce the same product/intermediate stream yield. EIA analysts developed a set (or slate) of crude oil blends from which refineries of each type are assumed to select.

4.7 Supply and demand curves

For each Conversion region within the LP, the Conversion Model develops isoelastic supply and demand curves for each of the four crude oil types that make up the crude oil blends and for the petroleum products represented in the LP. The supply and demand curves are based on a set of price/quantity pairs provided by the Logistics Model, along with supply and demand elasticities (estimated by analyst judgement) to shape the curves. The Conversion Model sends equilibrium regional quantities and prices of crude oil and petroleum products from these supply and demand curves to the Logistics Model.

4.8 Capacity expansion mode versus dispatch mode

The Conversion Model projections published in the IEO or provided to other GHySMo models are produced in two steps. In the first step, the Conversion Model is run in capacity expansion mode (CAPM) to establish the refinery capacity expansion requirements needed to balance production projections based on crude oil supply curves and petroleum product demand curves. In the second step, the Conversion Model is run in dispatch mode (DISP) to determine optimal economic refinery processing levels and prices given the same crude oil and product curves used in the CAPM mode and to incorporate the newly established refinery unit capacities instead of making its own build decisions. The LP has these two different modes because build decisions are made based on investment periods, and dispatch decisions are based on current year markets. So, the design of the time horizon differs between these two modes.

In CAPM mode, the model considers several planning horizons, each consisting of six planning periods. The first three planning periods in each horizon represent a single year, in which ADU and downstream upgrading unit capacity expansion may be projected. The remaining three planning periods in each horizon represent an aggregate of three years each (nine years in all), in which no additional capacity builds are projected. Any capacity expansion that is projected for the first three planning periods is included as planned capacity for all subsequent planning periods and planning horizons. The model then moves the starting model year forward three years for each of the subsequent planning horizons, until the last projection year is included in the first three planning periods of the planning horizon.

After the CAPM run, the Conversion Model is run again in DISP mode. In each DISP run, the model assumes a single planning horizon, and each planning period represents a single year through the end of the projection period. Although the DISP run is designed to project no additional capacity builds, it incorporates the expansions projected in the CAPM run as planned capacity additions.

4.9 Modeling platform for the Conversion Model

The Conversion Model's LP model structure and results are developed, executed, and examined using the AIMMS platform and modeling language and are stored as an AIMMS project (GHSM.aimms). It communicates with the rest of GHySMo using Python. The Conversion Model can be run integrated with GHySMo and WEPS, or stand-alone for development, testing, and diagnostics (from within the AIMMS developer).

4.10 Abbreviations and acronyms for the Conversion Model

AIMMS: Advanced Interactive Multidimensional Modeling System

CBOB: Conventional blendstock for oxygenate blending to produce finished gasoline for final sale

GHySMo: Global Hydrocarbon Supply Model

GRTMPS: Generalized Refinery, Transportation, and Marketing Planning System

Kerojet: Kerosene-based jet fuel

Petroleum_oth: Other petroleum products includes aviation gasoline, naphtha and other oils for feedstocks, and miscellaneous products

Petroleum_SEQ: Sequestered petroleum products includes asphalt and road oil, lubricants, and waxes

Resid: Residual fuel oil

4.11 Relationship to GHySMo and WEPS

Information flow into and out of the Conversion Model is as follows:

1. Logistics provides the following to Conversion Model by Conversion regions

- Price/quantity pairs for petroleum product demands
- Price/quantity pairs for crude oil supplies
- Biofuels demands (combined ethanol, biodiesel, and other)

2. Conversion Model provides the following to Logistics Model by Conversion regions

- Price/quantity pairs for production of petroleum products
- Diesel, kerojet, heating oil, CBOB, gasoline, other petroleum (naphtha/petchem), other sequestered byproducts (BTX, marketable coke), and LPG
- Price/quantity pairs for demands of crude oil supplied

- API gravity: light, medium, heavy, and condensate
- Biofuels supply price
- ethanol, biodiesel
- Demand/consumption of still gas and catalyst coke (fuel consumption)
- Production of sulfur at refineries
- Utility demands at refineries (natural gas, electricity, and steam)

3. Upstream Model interacts with Logistics Model to provide crude oil prices and supplies by Conversion regions, sent from the Logistics Model to the Conversion Model

Information flow into and out of the GHySMo Conversion Model from WEPS is as follows:

- Demands for petroleum products (including biofuels blends)
- Gross domestic product (GDP) factors

5. Logistics Model

5.1 Introduction

The GHySMo Logistics Model determines how to balance global hydrocarbon production and consumption for the IEO. The other GHySMo models (Upstream, Conversion, and Integration) provide the Logistics Model with lists of locations that have demands for commodities (crude oil types, natural gas, several refined petroleum products, etc.) that must be met or supplies of commodities that are available to meet these demands. The Logistics Model combines these lists into a single list of locations and then identifies optimal transportation routes between the supply and demand locations, using a database of existing, planned, and potential pipelines and shipping routes.[5] Using other available data, the model also identifies associated hydrocarbon product processing facilities[6] and shipping fleets.

The GHySMo Integration Model provides the Logistics Model with initial vectors of quantities of products supplied, quantities of products demanded, and the corresponding prices and price elasticities for each commodity at each location for the most recent historical year for which data are available and for each year in the projection period.

The Logistics Model first develops a *base network topology*, which consists of the following:

- All supply and demand locations
- All existing and potential transportation routes between the supply and demand locations
- The existing and planned capacities of all transportation routes (infinite for shipping routes)
- The existing, planned, and potential liquefaction and regasification facilities with existing and planned capacities
- The existing and planned capacities of all ship fleets

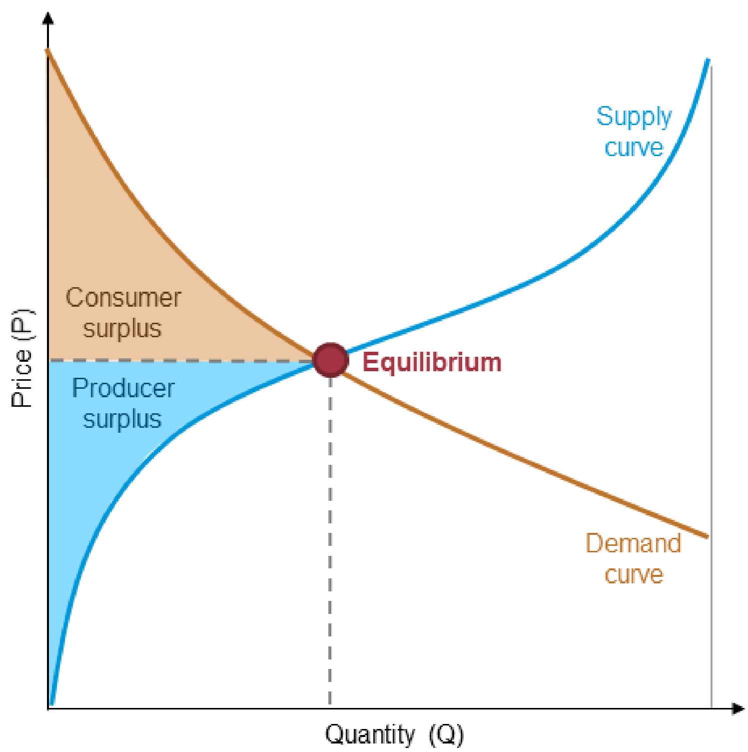
[5] As of *IEO2019*, the GHySMo Logistics Model does not consider rail or truck transport or storage of commodities.

[6] As of *IEO2019*, only liquefaction and regasification facilities for liquefied natural gas (LNG) shipping.

The model assumes that any transportation route/processing facility that represents only a *potential* connection or a *potential* liquefaction or regasification facility (in other words, not yet existing) has an initial capacity of zero.

The model uses a pair of linear programs (LPs) to balance supply and demand for hydrocarbon products. Each LP maximizes a representation of the total producer profits (revenues minus costs), or, equivalently (as both imply that a competitive market is in equilibrium), the sum of consumer and producer surplus across all commodities and all years of the projection period. The consumer surplus represents the amount of money saved by consumers who would buy a hydrocarbon product at a given price but are able to obtain it at a lower one. The producer surplus represents the added revenue of suppliers who could sell hydrocarbon products at a lower price but are able to charge a higher one.

Figure 3. Consumer and producer surplus



The Logistics Model provides the following outputs for use by the GHySMo Integration Model:

- Supplies of commodities from supply locations (in other words, crude oils, natural gas, and HGLs from upstream suppliers; HGLs and other petroleum products from refining locations)
- Demands for commodities at consumer locations
- Use, if any, of various commodities along supply chains (for example, use of natural gas in pipelines to power compressors)
- Liquefaction and regasification capacity, utilization, and flow levels at processing locations
- Wholesale commodity prices

Because the Logistics Model produces medium- to long-term projections of hydrocarbon supplies and prices, it faces theoretical issues. One would expect that the transportation and processing network connecting the regions of the world would evolve during the projection period through changes to capacities. The model is therefore designed to represent such changes, with or without intervention by EIA analysts. To simulate a hypothetical scenario, a modeler can specify changes to the network, but the model can also independently simulate changes to the network. The simulated changes yield positive, risk-adjusted *net present-value* (NPV) of hypothetical capacity changes, subject to analyst-specified (although possibly unlimited) bounds on these changes. The NPV is the difference between the values of cash or product inflows and outflows over a period of time; it used in investment planning to assess the profitability of a planned capacity change.

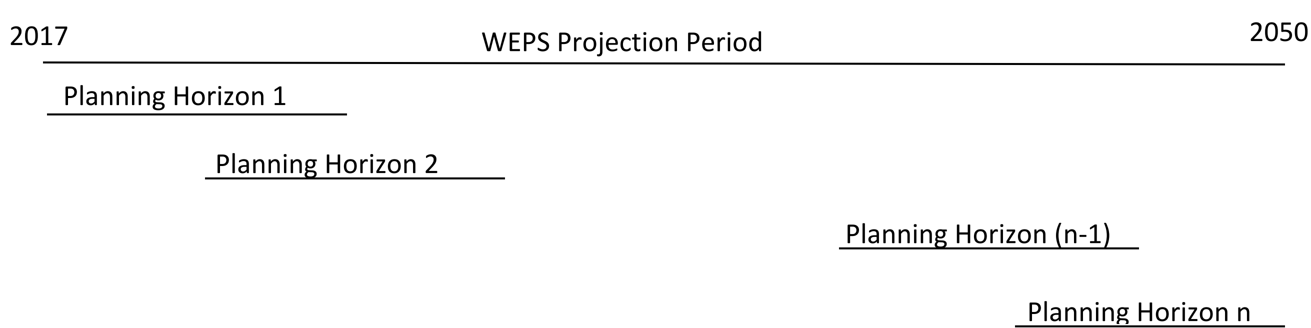
Enabling the model to simulate capacity expansion in the LP adds two complications:

- Allowing the model to maximize NPV during the entire projection period with perfect foresight deviates significantly from reality in which foresight is limited or non-existent and capacity change decisions must be continually reevaluated in light of new information.
- When using an LP model, the marginal price of a commodity at a node is represented by the value of the dual variable associated with the balancing constraint for that commodity at the specified location. Because the capacity change decision is based on costs that include the costs associated with the change, such a dual variable represents a long-run marginal cost. However, according to economic theory, decisions regarding the flow of commodities during each time period represented by the Logistics Model should be based on the short-run marginal costs.

To reconcile the requirement to represent capacity changes with decisions based on short-run marginal pricing, the Logistics Model incorporates two related LPs that model decisions during two different time horizons.

The first LP is referred to as the Capital Allocation and Planning Model (CAPM), and it maximizes the risk-adjusted NPV of total profits over a series of configurable multi-year planning horizons. These planning horizons are intended to be shorter than the total model projection period, as seen in Figure 2. For each planning horizon, the CAPM LP determines how to meet the demands from the available supplies, not only by choosing how much of each supply to use from each supply location and the route by which these supplies reach consumers, but also by changing the capacities of the pipeline routes, processing facilities, and shipping fleets.

Figure 4. CAPM LP planning horizons



The second LP is the dispatch (DISP) LP, and it maximizes total profits for a single year. After the CAPM LP runs once, projecting over an entire planning horizon and potentially updating capacities in the network in that horizon, the DISP LP takes the projected capacities as given and, for each year in the planning horizon, projects supplies, flows, and short-run marginal prices in the specified year.

Once the DISP LP computes projections for each of the years in a planning horizon, the CAPM LP projects for the next planning horizon, which is offset from the previous horizon by several years, as seen in Figure 2. The DISP LP then projects for each year in the new horizon, and then this sequence repeats until the DISP LP projects for the final year in the projection period.

Because the CAPM LP must also project how all the consumer demands will be met, given the supplies and possible flows, all of the constraints of the DISP LP are also included in the CAPM LP.

5.2 Objective functions

Both the CAPM and DISP LPs maximize projected profits. The objective functions represent sums of revenues obtained from selling commodities at their destinations *minus* all costs associated with selling the commodities. The cost factors are split into three major categories:

- Costs incurred at the demand, supply, and processing locations
- Costs incurred on transportation routes
- Costs incurred for the use of shipping fleets in general (but not associated with shipping between any specific locations)

The major difference between the two objective functions is how they account for time and risk. For the DISP LP, each of the three cost categories is also split into fixed and variable costs, and the variable costs are generally further split into costs that are incurred regardless of which commodity within a class (for example, *clean* liquids, such as gasoline or diesel; *dirty* liquids, such as crude oil or residual fuel oil; etc.) is being transported (for example, fuel use that depends only on the amount of the commodity transported, not on any other property of the commodity) and costs that are specific to the commodity being transported (for example, additional cleaning that may be necessary after transporting a more viscous product such as heavy crude oil). For the CAPM LP, there are capital costs in addition to the fixed and variable costs.

The objective function for the DISP LP is the sum of the revenues minus the sum of all of these cost components taken over all commodities and all locations for a single year.

Because the CAPM LP maximizes risk-adjusted net present values, the revenues and costs are adjusted. The revenues for each year and location are divided by the risk factor for the year and location. All costs are multiplied by the corresponding risk factor for the projection year. Each cost and revenue value is divided by a discount factor for the projection year.

5.3 LP decision variables

Decision variables are the unknown values for which an LP solves. For the DISP LP, each projection period has two sets of decision variables:

- The amount of each commodity supplied at each location
- The amount of each commodity that flows through each transportation route

All other decision variables defined for the DISP LP are defined as functions of variables in these two sets.

The CAPM LP starts with the same two sets of decision variables as the DISP LP, but it includes one additional set of decision variables for each period in the planning horizon: for each asset that can change capacity (for example, liquefaction facilities, pipelines, ship fleets), the LP solves for the amount of capacity (if any) to add to the existing capacity. Although the CAPM LP doesn't automatically simulate the removal of capacity from the network, EIA analysts can specify capacity decreases, resulting from LNG contracts, for liquefaction and regasification facilities.

5.4 Key constraints

The profit maximization objective is subject to constraints that impose physical, operational, and policy restrictions on the costs and revenues.

The following table shows the constraints and the LP(s) that they apply to:

Table 2. Key constraints

Constraint	CAPM	DISP
Constant local commodity stocks	X	X
Minimum and maximum capacity utilization factors	X	X
Minimum and maximum capacity increases	X	
Minimum and maximum capacity levels	X	

Constant local commodity stocks

At each location in each projection period, for each commodity, the total mass of the commodity added by local supply or net deliveries received (gross deliveries less any losses in unloading) must equal the sum of the mass of commodity removed from local demand and gross shipments to other locations (including any losses in loading). In other words, the model assumes that no stocks are built or depleted over time at any location. In the DISP LP, this constraint determines the marginal (shadow) price of each commodity at each location that is ultimately returned to the GHySMo Integration Model.

Minimum and maximum capacity utilization factors

For each time period, the model assigns to each item that has a finite capacity (liquefaction and regasification facilities, pipelines, and ship fleets) a minimum and maximum capacity utilization factor that can vary from zero to one. A utilization factor of zero indicates that the item is not being used, while a factor of one indicates that it's being used to its full capacity. The model, by default, sets the minimum utilization factors to zero and the maximum utilization factors to one. EIA analysts, however, adjust the minimum and maximum capacity utilization factors to account for known or expected future events, for example, setting a maximum utilization factor to a value less than one to account for an equipment maintenance period.

Minimum and maximum capacity increases

In the CAPM LP, in each period, each asset (processing facility or pipeline transportation route) with a finite capacity has minimum and maximum capacity increase constraints. The model sets the minimum capacity increase to zero and the maximum capacity increase to infinity. So, the model *can* solve for any amount of capacity to add in a given period. EIA analysts, however, adjust the capacity increase bounds to account for known or expected future changes. For example, a project with a known capacity may be under construction and scheduled for completion at a known time. EIA analysts then set the minimum and maximum capacity increase to each be equal to this known capacity increment in the completion year. Similarly, it may be known that no new capacity will be added at a location for the first few years of the projection period, in which case the minimum and maximum capacity increase values should each be set to zero for those years.

Minimum and maximum capacity levels

The analyst-specified capacity increase constraints allow EIA analysts to account for known future changes in capacity. The minimum and maximum capacity level constraints are intended to provide an alternate way accounting for future changes when less information is available. EIA analysts may, for example, be aware of capacity being built in a specific location but may not have information about the year-by-year build plans. For example, they may know of a plan to increase liquefaction capacity for several years at a location, without knowing the yearly capacity increases. By specifying a minimum capacity level several years in the future, rather than specifying yearly minimum capacity increases, analysts can let the model solve for the build plan that makes the most economic sense, while ensuring that the model represents the full increased capacity in the location by the expected completion year.

5.5 Interaction with other GHySMo and WEPS Models

From the Upstream Model and the Conversion Model, the Logistics Model receives a list of hydrocarbon producing and processing locations. From the GHySMo Integration Model, the Logistics Model receives the list of WEPS demand regions. Using a combination of analyst judgement, a database of known existing and under-construction pipelines, a database of shipping routes, and Dijkstra’s shortest-path algorithm, the Logistics Model identifies pipeline and shipping routes between supply and demand locations.

From the Upstream Model (by way of the Integration Model), the Logistics Model receives a list of projected base prices and quantities and own-price supply elasticities for each of the commodities that each of the supply locations are capable of supplying. The Logistics Model uses the base price and quantity and the elasticity to generate an isoelastic supply curve for each commodity for supply location. Also by way of the GHySMo Integration Model, the Logistics Model receives aggregate demands for finished petroleum products, natural gas, and HGLs for each of the WEPS demand regions.

From the Conversion Model, by way of the Integration Model, the Logistics Model receives two sets of data for each of the refining locations. First, it receives the demands for each of the types of crude oils and natural gas. Second, it receives base prices and quantities and own-price supply elasticities for each of the finished petroleum products that the refining locations are capable of supplying.

Upon successful completion, the Logistics Model returns to the Integration Model a spreadsheet with the outputs described in the Introduction.

5.6 Model data sources

Existing and planned infrastructure

Existing liquefaction and regasification capacities and capacities of projects under construction were aggregated from International Gas Union (IGU) data.

Existing pipeline capacities and capacities of pipeline projects under construction were aggregated from IHS EDIN.

Ship fleet sizes were aggregated from several data sources, including the Iberia Capital Partners November 14, 2011, *Coal and Shipping Industry Report* and the International Group of Liquefied Natural Gas Importers (GIIGNL) 2016 Annual Report.

Capital costs

Capital costs for liquefaction and regasification plants were estimated from IHS EDIN data as a cost per unit capacity in different regions.

Operations & maintenance (O&M) costs

Variable O&M costs (in other words, costs that are a function of the amount of commodity that flows) on pipelines were estimated by linear regression from Federal Energy Regulatory Commission (FERC) data for a sample of crude oil transporting pipelines. The average dollar per mile (per metric ton per year) O&M costs were estimated using data from FERC filings for a sample of pipelines. Multiplying the O&M rate by the length of the aggregate pipelines (using IHS EDIN data) gave a variable O&M cost per pipeline.

Translating from U.S. costs to international costs

The ratio of a country’s per-capita GDP to the U.S. per-capita GDP was used to translate all costs from U.S. dollars to a commensurate cost in the target country.